OECD Reviews of Evaluation and Assessment in Education

Synergies for Better Learning

AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT



OECD

OECD Reviews of Evaluation and Assessment in Education

Synergies for Better Learning

AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT



This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

ISBN 978-92-64-19064-1 (print)
ISBN 978-92-64-19065-8 (PDF)
http://dx.doi.org/10.1787/9789264190658-en

Series: OECD Reviews of Evaluation and Assessment in Education
ISSN 2223-0947 (print)
ISSN 2223-0955 (online)

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Photo credits: Cover © iQoncept - Fotolia.com, © AKS - Fotolia.com, © Sergej Khackimullin - Fotolia.com.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

© OECD 2013

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of the source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org*. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) at *contact@cfcopies.com*.

Foreword

Authentic, valid and reliable evaluation and assessment, those which lead to the improvement of educational practices at all levels and lift student learning, are central to establishing a high-performing education system. They are also instrumental in recognising and rewarding the work of educational practitioners and in certifying the learning of students. Promoting evaluation and assessment is clearly in the interest of students and their families, educational practitioners and education systems. As a result, more and more countries embark on ambitious school reform programmes which include a strong element of evaluation and assessment. This can consolidate the evaluation culture in education systems and reinforces the role of evaluation and assessment frameworks in driving the reform agenda.

In late 2009, the OECD Education Policy Committee embarked on a comprehensive international review of evaluation and assessment policies, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Its goal was to provide analysis and policy advice to countries on how evaluation and assessment arrangements can be embedded within a consistent framework that can bring about real gains in performance across the school system. In addition to this publication, the Review generated 25 reports by participating countries, 15 reports by external review teams (released as a publication series, *OECD Reviews of Evaluation and Assessment in Education*) and several research papers (all available on the OECD website at *www.oecd.org/edu/evaluationpolicy*). This OECD Review provides an unparalleled comprehensive analysis of evaluation and assessment policy issues at international level.

OECD work helps countries to learn from one another. It can also highlight issues and explore policy options that may be difficult to raise in national debates. Both of these elements clearly underpin this report and the work behind it. The direct involvement of countries has been crucial to the process. The 28 countries that were actively engaged committed substantial resources and opened their evaluation and assessment policies to external review and debate. This collaborative approach enabled countries to learn more about themselves and to add to the broader knowledge base by sharing evidence on the impact of policy reforms and the circumstances under which they work best.

The Review was overseen by the OECD Group of National Experts (GNE) on Evaluation and Assessment, which was established as a subsidiary body of the Education Policy Committee. The GNE on Evaluation and Assessment was chaired by Mr. Gábor Halász, Professor, Eötvös Loránd University, Hungary, and had as Vice-Chairs Mr. Éamonn Murtagh, Assistant Chief Inspector, Department of Education and Skills, Ireland, and Mr. Kwangho Kim, Research Fellow, Korean Educational Development Institute, Korea. We are grateful for their insightful and committed guidance.

The Review benefited substantially from the involvement of organisations representing students, parents, teachers, school leaders, schools, teacher educators, researchers and employers. Their representatives served on national committees, prepared written submissions, met with review teams and participated in conferences and meetings of the GNE on Evaluation and Assessment. The Review also benefited from the contribution of the Business and Industry Advisory Committee to the OECD (BIAC) and the Trade Union Advisory Committee to the OECD (TUAC), as well as other international organisations interested in evaluation and assessment policy, including the European Commission, Eurydice, the Standing International Conference of Inspectorates (SICI), UNESCO and the World Bank.

Annex A details the many people and organisations who contributed to the Review as national co-ordinators, members of the GNE on Evaluation and Assessment, members of country review teams, and authors of country background reports and commissioned research papers – more than 120 people in all. We are fortunate to have benefited from their expertise and are appreciative of their commitment to the Review. The 15 review visits involved 30 external reviewers with a range of research and policy backgrounds. Overall, the external review teams visited about 90 schools and met with about 2 800 individuals on which the country review findings were based. We thank them all for their valuable contributions to the collective knowledge base.

Within the Directorate for Education and Skills, the Review was carried out by the Education and Training Policy Division under the leadership of Deborah Roseveare (from October 2009 until January 2012) and by the Early Childhood and Schools Division under the leadership of Michael Davidson (from February 2012 on). Within the OECD Secretariat, Paulo Santiago (co-ordinator), Deborah Nusche and Claire Shewbridge were responsible for the Review and the preparation of this report. They also took responsibility for individual country reviews. Heike-Daniela Herzog took responsibility for the administrative work within the Review, the organisation of meetings and communication with the countries. A number of other colleagues contributed to both the Review and this report (see Acknowledgements below). A larger group of colleagues within the OECD provided advice at key stages. In particular, collaboration was established with the Programme for International Student Assessment (PISA), the Teaching and Learning International Survey (TALIS), CERI's Innovative Learning Environments (ILE) project, the INES Network for the Collection and the Adjudication of System-Level Descriptive Information on Educational Structures, Policies and Practices (NESLI), CERI's work on Longitudinal Information Systems as part of its Innovation Strategy for Education and Training, CERI's work on Governing Complex Education Systems and the Centre for Effective Learning Environments' (CELE) work on evaluating quality in educational facilities.

This report was released in Oslo on 11 April 2013 at an International Conference jointly sponsored by the Ministry of Education and Research of Norway, the Norwegian Directorate for Education and Training, and the OECD.

The OECD intends to maintain the momentum of its work on evaluation and assessment policies and to build on the Review on Evaluation and Assessment Frameworks for Improving School Outcomes and this report.

ACKNOWLEDGEMENTS

The authors of this report are Deborah Nusche, Thomas Radinger, Paulo Santiago (co-ordinator) and Claire Shewbridge of the Directorate for Education and Skills, with the following distribution of responsibilities as lead authors:

Chapter 1. The focus on evaluation and assessment	Paulo Santiago
Chapter 2. Trends in evaluation and assessment	Paulo Santiago
Chapter 3. The evaluation and assessment framework	Paulo Santiago
Chapter 4. Student assessment	Deborah Nusche
Chapter 5. Teacher appraisal	Deborah Nusche
Chapter 6. School evaluation	Claire Shewbridge
Chapter 7. The appraisal of school leaders	Thomas Radinger
Chapter 8. Education system evaluation	Claire Shewbridge

Heike-Daniela Herzog was responsible for the production and layout of the report and contributed to its editing and proofreading. Valuable comments on draft chapters were provided by members of the OECD Secretariat, members of the Group of National Experts on Evaluation and Assessment, members of the OECD Education Policy Committee, national co-ordinators of participating countries, and researchers and international agencies associated with the work. The team of authors is grateful to individual experts who contributed to the country-specific *OECD Reviews of Evaluation and Assessment in Education*, whose expertise and analysis have fed into this report (they are listed in Table A.3 of Annex A).

Thanks are due to the many people who worked on this Review at different stages of its development, in addition to its core members (Heike-Daniela Herzog, Deborah Nusche, Paulo Santiago, Claire Shewbridge): Morten Rosenkvist (February to July 2010) on secondment from the Ministry of Education and Research of Norway; Stefanie Dufaux (September 2010 to August 2011) as a Fellow of the Carlo Schmid Programme and as a Consultant; and Thomas Radinger (September 2011 to September 2012) as a Fellow of the Carlo Schmid Programme and as a Consultant. They provided substantial input into the Review's knowledge base through their analytical work. Thomas Radinger also assumed responsibility for the analysis on school leader appraisal and co-ordinated the preparation of the tables summarising features of evaluation and assessment frameworks. Both Stefanie Dufaux and Thomas Radinger took responsibility for statistical work within the Review. Violaine Faubert, Marlène Isoré, Allison Morris and Anne Nayral de Puybusque provided research assistance summarising key areas of the literature on evaluation and assessment during their internships at the OECD. Janet Looney, as an external Consultant, prepared research papers as background analytical work for the Review. Francisco Benavides and Diana Toledo, of the OECD Secretariat, contributed as experts to the OECD Reviews of Evaluation and Assessment in Chile and Mexico, respectively.

Sabrina Leonarduzzi (from June until December 2009) and Sara-Jayne Moss (from January until June 2010) provided administrative support for the Review. Cassandra Davis (Directorate for Education and Skills) and Anne-Lise Prigent (Public Affairs and Communications Directorate) provided advice on the production of the report and dissemination activities. Brian Keeley (Public Affairs and Communications Directorate) prepared the executive summary of this report. Therese Walsh and Damian Garnys (both from the Public Affairs and Communications Directorate) provided editorial assistance.

We are also grateful to the German Academic Exchange Service (DAAD) and the German National Academic Foundation for sponsoring Stefanie Dufaux's (September 2010 – March 2011) and Thomas Radinger's (September 2011 – June 2012) appointments with the OECD Secretariat as Fellows of the Carlo Schmid Programme. Gratitude is also extended to the Ministry of Education and Research of Norway for supporting the secondment of Morten Rosenkvist with the OECD Secretariat and for hosting the International Conference to launch this report (11-12 April 2013).

This book has...



StatLinks

A service that delivers Excel® files from the printed page!

Look for the *StatLinks* at the bottom right-hand corner of the tables or graphs in this book.
To download the matching Excel® spreadsheet, just type the link into your Internet browser, starting with the ***http://dx.doi.org*** prefix.
If you're reading the PDF e-book edition, and your PC is connected to the Internet, simply click on the link. You'll find *StatLinks* appearing in more OECD books.

Table of contents

Executive summary **13**

Overview **17**

Chapter 1. **The focus on evaluation and assessment** **29**

The growing importance of evaluation and assessment 30
Methodology 30
Organisation of the report 32

Chapter 2. **Trends in evaluation and assessment** **35**

Introduction 36
Trends in evaluation and assessment 36
Contextual developments shaping evaluation and assessment 45
References 53

Chapter 3. **The evaluation and assessment framework: Embracing a holistic approach** **57**

Introduction 58
Definitions and scope 59
Educational context 61
Governance 62
Overarching reference: Goals for student learning 73
Design 78
Capacity for evaluation and assessment 97
Use of results 102
Implementation of evaluation and assessment policies 108
Pointers for future policy development 117
References 129

Chapter 4. **Student assessment: Putting the learner at the centre** **139**

Introduction 140
Analytical approach 140
Impact, drivers and contextual developments 144
Governance 152
Procedures 170
Capacity 191
Reporting and use of results 202
Pointers for future policy development 213
References 225
Annex 4.A1 Features of student assessment frameworks in lower secondary education 239
Annex 4.A2 Features of student assessment frameworks in primary and upper secondary education 269

Chapter 5. Teacher appraisal: Enhancing teacher professionalism 271

Introduction 272
Analytical approach 272
Impact, drivers and contextual developments 274
Governance 280
Procedures 292
Capacity 312
Use of results 322
Pointers for future policy development 332
References 343
Annex 5.A Features of teacher appraisal frameworks 353

Chapter 6. School evaluation: From compliancy to quality 383

Introduction 384
Analytical approach 384
Impact, drivers and contextual developments 386
Governance 393
Procedures 414
Capacity 432
Reporting and use of results 445
Pointers for future policy development 463
References 473

Chapter 7. The appraisal of school leaders: Fostering pedagogical leadership in schools 485

Introduction 486
Analytical approach 486
Impact, drivers and contextual developments 496
Governance 506
Procedures 517
Capacity 539
Use of results 543
Pointers for future policy development 550
References 561
Annex 7.A Features of school leader appraisal frameworks 569

Chapter 8. Education system evaluation: Informing policies for system improvement 583

Introduction 584
Analytical approach 584
Impact, drivers and contextual developments 587
Governance 590
Procedures 595
Capacity 618
Reporting and use of results 626
Pointers for future policy development 639
References 649

Annex A. How the Review on Evaluation and Assessment Frameworks for Improving School Outcomes was conducted 657

Annex B. Summary of policy directions from the Review on Evaluation and Assessment Frameworks for Improving School Outcomes 667

Tables

Table 3.1 Use of results for accountability and development across countries (2012) 64
Table 3.2 Specialised intermediate agencies with a role in the evaluation and assessment framework, by country (2012) 69
Table 3.3 Comprehensiveness and degree of structure of evaluation and assessment frameworks across countries (2012) 89
Table 3.4 Key articulations within components of the evaluation and assessment framework 93
Table 3.5 Key articulations between components of the evaluation and assessment framework 95
Table 4.1 References used in student assessment in lower secondary education – ISCED 2 (2012) 169
Table 4.2 Characteristics of central examinations in lower secondary education – ISCED 2 (2011-12) 173
Table 5.1 Responsibilities for determining procedures for teacher appraisal (2011-12) 291
Table 5.2 Reference documents used for different types of teacher appraisal (2011-12) 295
Table 5.3 Evaluators involved in teacher appraisal processes (2011-12) 312
Table 5.4 Evaluators involved in teacher appraisal processes (2011-12) 319
Table 5.5 Influence of teacher appraisal on professional development (2011-12) 323
Table 5.6 Influence of teacher appraisal on career decisions (2011-12) 326
Table 5.7 Rewards connected to teacher appraisal (2011-12) 329
Table 5.8 Responses to underperformance of teachers (2011-12) 331
Table 6.1 Purpose of school evaluation 394
Table 6.2 Responsibilities for external school evaluation 401
Table 6.3 Legal frameworks for external school evaluation 403
Table 6.4 Requirements for school self-evaluation 406
Table 6.5 Frequency of school self-evaluations 409
Table 6.6 Links between external school evaluation and school self-evaluation (2009) 412
Table 6.7 Areas addressed during school evaluation 415
Table 6.8 Collecting information for compliance-related reporting 425
Table 6.9 Classroom observation indicators to evaluate the quality of teaching and learning 428
Table 6.10 National initiatives to support school self-evaluation 437
Table 6.11 Use of school evaluation results and school performance measures for accountability (2009) 447
Table 6.12 Use of the results of external school evaluation for accountability (2009) 454
Table 7.1 A taxonomy of country approaches to the appraisal of school principals in public schools 507
Table 7.2 The use of school leader appraisal results 544
Table 8.1 Policy frameworks and development of education system evaluation (2012) 592
Table 8.2 Indicators of a strategic approach to education system evaluation in OECD countries (2012) 593
Table 8.3 National assessments in OECD countries (2012) 601
Table 8.4 Testing formats used in national assessments (2012) 607
Table 8.5 OECD country participation in international achievement studies (2012) 609
Table 8.6 The use of stakeholder surveys in the OECD (2012) 612
Table 8.7 Availability of longitudinal information for education system evaluation (2012) 615
Table 8.8 Capacity for evaluating the education system and sub-national systems (2012) 619
Table 8.9 National reporting on the results of education system evaluation (2012) 627

Figures

Figure 3.1 Main features of the overall evaluation and assessment framework 60
Figure 3.2 Student test scores as a criterion of teacher appraisal and school evaluation (2007-08) 80
Figure 3.3 Publication of school evaluations (2007-08) 82
Figure 3.4 Publication of student achievement data (2009) 83
Figure 3.5 Direct appraisal of classroom teaching as a criterion of teacher appraisal and school evaluation (2007-08) 84
Figure 3.6 Participation rates in professional development activities involving the sharing of classroom practice (2007-08) 85
Figure 3.7 Teachers who received no appraisal or feedback and teachers in schools that had no school evaluation in the previous five years (2007-08) 88
Figure 3.8 Conceptual framework to analyse evaluation and assessment in school systems 91
Figure 3.9 Articulations within the evaluation and assessment framework 92
Figure 3.10 Teachers' perceptions of professional development needs in student assessment practices (2007-08) 99
Figure 3.11 School principals according to their management styles (2007-08) 101
Figure 4.1 Conceptual framework for student assessment 143
Figure 4.2 Existence of central frameworks for internal summative assessment at ISCED levels 1-3 (2012) 154
Figure 4.3 Existence of standardised central examinations at ISCED levels 1-3 (2012) 155
Figure 4.4 Existence of central frameworks for formative assessment at ISCED levels 1-3 (2012) 156
Figure 4.5 Existence of standardised central assessments with no stakes for students at ISCED levels 1-3 (2012) 158
Figure 5.1 Conceptual framework for teacher appraisal 275
Figure 5.2 Teacher perceptions of the appraisal and/or feedback they received (2007-08) 276
Figure 5.3 Existence of policy frameworks for teacher appraisal across countries (2011-12) 282
Figure 5.4 Teachers missing out on appraisal and feedback opportunities (2007-08) 288
Figure 5.5 Direct appraisal of classroom teaching as an aspect of teacher appraisal (2007-08) 301
Figure 5.6 Student test results as an aspect of teacher appraisal (2007-08) 305
Figure 5.7 Student and parent feedback as an aspect of teacher appraisal (2007-08) 308
Figure 6.1 Conceptual framework for school evaluation 387
Figure 6.2 Evidence on reported improvement actions in schools following external school evaluation 392
Figure 6.3 Criteria for student learning outcomes used in school evaluation (2007-08) 423
Figure 6.4 Direct appraisal of classroom teaching in school evaluation 426
Figure 6.5 School use of standardised tests and assessment data (PISA 2009) 429
Figure 6.6 Use of student assessment results in school self-evaluation (PISA 2009) 451
Figure 6.7 Public reporting of student performance and reporting to parents (PISA 2009) 460
Figure 7.1 Conceptual framework for school leadership appraisal 495
Figure 7.2 School autonomy over resource allocation (2009) 503
Figure 7.3 School autonomy over curricula and assessment (2009) 504
Figure 7.4 School principals observing instruction in classrooms (2009) 526
Figure 7.5 School principals using exam results for curriculum development (2009) 527
Figure 7.6 Use of student achievement data in the appraisal of school principals' performance (2009) 538
Figure 8.1 Conceptual framework for education system evaluation 586
Figure 8.2 School monitoring by administrative authorities (PISA 2009) 625

Boxes

Box 3.1 The Comprehensive Reform of Basic Education in Mexico 77
Box 3.2 Availability of evaluation experts in Denmark and Hungary 100
Box 3.3 Data information systems in Australia (New South Wales) 104
Box 3.4 Collaboration initiatives and partnerships in Norway 105
Box 3.5 Promoting dialogue and reaching common views on educational evaluation in Denmark 111
Box 4.1 How student assessment influences learning outcomes: A brief overview of research evidence 145
Box 4.2 Key competencies around the world 148
Box 4.3 Matching a commitment to formative assessment with concrete support for teachers 157
Box 4.4 Defining and communicating the purposes of assessment 161
Box 4.5 Monitoring student learning in the Netherlands 165
Box 4.6 Classification of learning outcomes: A brief overview 171
Box 4.7 Assessment formats that measure broader competencies 178
Box 4.8 Continuous teacher-based assessment of student competencies 180
Box 4.9 Singapore: Creating "thinking schools" 182
Box 4.10 Sophisticated ICT-based assessments 186
Box 4.11 Canada: Policies on reporting student performance 206
Box 4.12 Innovations in the reporting system in Australia 207
Box 4.13 Organising remedial education and student orientation based on formative assessment results 212
Box 5.1 How teacher appraisal influences teaching and learning: A brief overview of research evidence 277
Box 5.2 Research and experiences regarding probation 284
Box 5.3 Appraisal systems for registration and certification in Australia and the United States 285
Box 5.4 Using appraisal results for performance review and professional development 286
Box 5.5 Involving the teaching profession in the development of professional standards 297
Box 5.6 The Good Teaching Framework in Chile 299
Box 5.7 United States: Using evidence of student growth to appraise teachers 307
Box 5.8 Using student feedback as formative feedback for teachers 309
Box 5.9 Using multiple instruments for teacher appraisal 311
Box 5.10 Developing central expertise for teacher appraisal in Chile 314
Box 5.11 Supporting school leaders in developing skills for teacher appraisal 316
Box 5.12 Building capacity for peer appraisal 318
Box 5.13 Linking teacher appraisal to professional development 324
Box 5.14 Advanced Skills Teaching positions in Australia 327
Box 5.15 Singapore: Linking teacher appraisal to career pathways 328
Box 6.1 Strengthening the focus on school development 395
Box 6.2 Strengthening the focus on school accountability 396
Box 6.3 Self-review at the heart of school evaluation: New Zealand 413
Box 6.4 Evaluating a school's capacity for improvement 416
Box 6.5 Example of areas addressed in school evaluation: Canada and the Flemish Community of Belgium 418
Box 6.6 Indicators used to judge school quality: The Netherlands 420
Box 6.7 Developing evaluation criteria for school self-evaluation 421
Box 6.8 Collecting feedback from students, teachers and parents 430
Box 6.9 Recruiting senior educators to join external school evaluation teams, Northern Ireland, United Kingdom 433
Box 6.10 Effective self-evaluation 436
Box 6.11 Centrally developed tools for self-evaluation in Scotland, the United Kingdom 439

Box 6.12 School self-evaluation related competencies for school leadership: Ontario, Canada 440
Box 6.13 Targeted training on school self-evaluation for school principals in New Zealand 441
Box 6.14 Building professional learning communities: A challenge shared 443
Box 6.15 Emerging peer reviews among schools in the Flemish Community of Belgium 445
Box 6.16 Using school self-evaluation results in school improvement planning 449
Box 6.17 Factors promoting use of assessment data for self-evaluation 452
Box 6.18 Follow up on implementation of external school evaluation recommendations: Ireland 456
Box 6.19 Reporting results of external school evaluation 457
Box 6.20 The public reporting of school performance 462
Box 7.1 Aligning the school evaluation process and the individual appraisal of school leaders: The School Accountability and Performance Improvement Framework in the Northern Territory, Australia 489
Box 7.2 The appraisal of distributed school leadership: The Appraisal of Collective Performance process (*Asignación de Desempeño Colectivo*) in Chile 491
Box 7.3 Embedding school leader appraisal within a comprehensive leadership framework and strategy: The cases of Victoria, Australia, and Ontario, Canada 501
Box 7.4 School leader appraisal procedures at a local level 510
Box 7.5 Properties of school leader appraisal 512
Box 7.6 Approaches to combining developmental and accountability functions within one appraisal framework 514
Box 7.7 The introduction of two separate appraisal schemes for different purposes: The cases of Chile and Korea 515
Box 7.8 Developing professional school leadership standards in federally governed education systems 530
Box 7.9 Different approaches to the definition of professional school leadership standards 531
Box 8.1 The impact of PISA on national policy making 590
Box 8.2 The influence of developments in public sector performance management 594
Box 8.3 Benchmarks for education systems: European Union 598
Box 8.4 The development of indicator frameworks for system evaluation in Australia and Canada 599
Box 8.5 Sample surveys in the Netherlands and New Zealand 604
Box 8.6 The major international surveys providing information to evaluate the education system 610
Box 8.7 Longitudinal research programmes in the Netherlands 614
Box 8.8 Tensions between monitoring changes over time and holding schools accountable 617
Box 8.9 National agencies with specific responsibility for education system evaluation 621
Box 8.10 Commissioning specific education reviews in Korea and the Netherlands 623
Box 8.11 Annual analytical reports on the education system: Czech Republic, Norway and Portugal 629
Box 8.12 Examples of reports on the education system 631
Box 8.13 An illustration of the intended use of the results of education system evaluation: Slovak Republic 632
Box 8.14 Strategy to make education system evaluation results more accessible in Sweden 634
Box 8.15 Promoting the use of results from international comparative studies in Slovenia 635
Box 8.16 Stakeholder discussion of major education system results in Denmark and the Flemish Community of Belgium 635
Box 8.17 Support for evidence-based policy making within ministries 637
Box 8.18 Information systems for evaluation of local school systems: The United States 639

Executive summary

Introduction

Governments and education policy makers are increasingly focused on the evaluation and assessment of students, teachers, school leaders, schools and education systems. These are used as tools for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school, school leadership and teaching practices.

Results from assessment and evaluation are becoming critical to establishing how well school systems are performing and for providing feedback, all with the goal of helping students to do better.

This report compares the experience of 28 OECD countries, analyses the strengths and weaknesses of different approaches, and offers policy advice on using evaluation and assessment to improve the quality, equity and efficiency of education. It draws on a major study, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

Common themes

Several factors are driving the increased use of evaluation and assessment, including:

- An increased demand for effectiveness, equity and quality in education to meet economic and social challenges.
- A trend in education towards greater school autonomy, which is fuelling a need to monitor how schools are doing.
- Improvements in information technology, which allow for the development of both large-scale and individualised student assessment and facilitate the sharing and management of data.
- Greater reliance on evaluation results for evidence-based decision making.

The current state and use of evaluation and assessment varies greatly between OECD countries, but there are common themes:

Evaluation is expanding and becoming more diverse

Most OECD countries now see evaluation and assessment as playing a central strategic role, and are expanding their use. They are also taking a more comprehensive approach: Formerly, evaluation and assessment focused mainly on student assessment, but the focus is now broader and includes greater use of external school evaluation, appraisal of teachers and school leaders, and expanded use of performance data.

Indicators are rising in importance

Education systems are placing a stronger focus on measuring student outcomes, allowing comparisons of performance between schools and regions and over time. Most countries now have national databases on education, and issue education statistics and indicators. International benchmarking is also increasingly common and is informing national education debates.

Results are being put to varied uses

Results are being used to identify where schools are performing well and where they may need to improve. They are also being used to hold policy makers, school leaders and teachers accountable. For example, many countries now publish national tables of school results for use by, among others, parents, government officials and the media.

Rising reliance on educational standards

Many countries now set educational standards for what students should know and what they should be able to do at different stages of the learning process. This has encouraged monitoring to determine if students are meeting these standards.

Challenges and directions

Countries have different traditions in evaluation and assessment and take different approaches. Nevertheless, there are some clear policy priorities:

Take a holistic approach

To achieve its full potential, the various components of assessment and evaluation should form a coherent whole. This can generate synergies between components, avoid duplication and prevent inconsistency of objectives.

Align evaluation and assessment with educational goals

Evaluation and assessment should serve and advance educational goals and student learning objectives. This involves aspects such as the alignment with the principles embedded in educational goals, designing fit-for-purpose evaluations and assessments, and ensuring a clear understanding of educational goals by school agents.

Focus on improving classroom practices

The point of evaluation and assessment is to improve classroom practice and student learning. With this in mind, all types of evaluation and assessment should have educational value and should have practical benefits for those who participate in them, especially students and teachers.

Avoid distortions

Because of their role in providing accountability, evaluation and assessment systems can distort how and what students are taught. For example, if teachers are judged largely on results from standardised student tests, they may "teach to the test", focusing solely on skills that are tested and giving less attention to students' wider developmental and

educational needs. It is important to minimise these unwanted side-effects by, for example, using a broader range of approaches to evaluate the performance of schools and teachers.

Put students at the centre

Because the fundamental purpose of evaluation and assessment is to improve student learning, students should be placed at the centre. They should be fully engaged with their learning and empowered to assess their own progress (which is also a key skill for lifelong learning). It is important, too, to monitor broader learning outcomes, including the development of critical thinking, social competencies, engagement with learning and overall well-being. These are not amenable to easy measurement, which is also true of the wide range of factors that shape student learning outcomes. Thus, performance measures should be broad, not narrow, drawing on both quantitative and qualitative data as well as high-quality analysis.

Build capacity at all levels

Creating an effective evaluation and assessment framework requires capacity development at all levels of the education system. For example, teachers may need training in the use of formative assessment, school officials may need to upgrade their skills in managing data, and principals – who often focus mainly on administrative tasks – may need to reinforce their pedagogical leadership skills. In addition, a centralised effort may be needed to develop a knowledge base, tools and guidelines to assist evaluation and assessment activities.

Manage local needs

Evaluation and assessment frameworks need to find the right balance between consistently implementing central education goals and adapting to the particular needs of regions, districts and schools. This can involve setting down national parameters, but allowing flexible approaches within these to meet local needs.

Design successfully, build consensus

To be designed successfully, evaluation and assessment frameworks should draw on informed policy diagnosis and best practice, which may require the use of pilots and experimentation. To be implemented successfully, a substantial effort should be made to build consensus among all stakeholders, who are more likely to accept change if they understand its rationale and potential usefulness.

Overview

The growing focus on evaluation and assessment

There is widespread recognition that evaluation and assessment arrangements are key to both improvement and accountability in school systems. This is reflected in their increasing importance in national education agendas. As countries strive to transform their educational systems to prepare all young people with the knowledge and skills needed to function in rapidly changing societies, some common policy trends can be observed in one form or another in most OECD countries, including decentralisation, school autonomy, greater accountability for outcomes and a greater knowledge management capacity. Decentralisation and school autonomy are creating a greater need for the evaluation of schools, school leaders and teachers while greater IT capacity allows for the development and analysis of large-scale student assessments as well as individualised assessment approaches. Results from evaluation and assessment are becoming critical to knowing whether the school system is delivering good performance and to providing feedback for further development. Evaluation and assessment are instrumental in defining strategies for improving practices within school systems with the ultimate goal of enhancing student outcomes. These developments are having a strong influence in the way in which policy makers monitor system, school, school leader, teacher and student performance.

Countries increasingly use a range of approaches for the evaluation and assessment of students, teachers, school leaders, schools and education systems. These are used as tools for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school, school leadership and teaching practices. Strong emphasis is being placed on better equipping and encouraging teachers to carry out self-appraisal and student formative assessment, on providing the incentives and means for school self-evaluation, on encouraging "value-added" evaluation and on more regular standardised testing of students and national monitoring of the overall system. However, countries often face difficulties in implementing evaluation and assessment procedures. These may arise as a result of poor policy design, lack of analysis of unintended consequences, little capacity for educational agents to put procedures into practice, lack of an evaluation culture, or deficient use of evaluation results.

This report is concerned with evaluation and assessment policies in school systems that can help countries achieve their educational goals and student learning objectives. It draws on a major study, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

Main trends within evaluation and assessment

Although not all countries are in the same position, a number of trends within evaluation and assessment emerge.

Educational evaluation in school systems is expanding

It is apparent that education policy is increasingly conferring a central strategic role to evaluation and assessment as indispensable tools for improvement, accountability, educational planning and policy development. In the last two decades, most countries have introduced a wide range of measures intended to improve evaluation and assessment at all levels from the student to the school system itself. The expansion of educational evaluation results from increased demands for effectiveness, equity and quality in education so new economic and social needs are met. The greater importance of evaluation and assessment in education policy has involved the creation of specifically dedicated agencies which assume a central role in the governance of the evaluation and assessment framework. This recognises the need for specialised expertise, the imperative of building adequate capacity to deliver evaluation and assessment policies and the necessity of introducing some independence *vis-à-vis* education authorities.

There is a greater variety of evaluation and assessment activities

The expansion of educational evaluation has been accompanied by considerable diversification of evaluation and assessment activities. Although educational evaluation within school systems is not a recent concern, it has traditionally focussed mostly on the assessment of students. In recent years, countries are increasingly developing more comprehensive evaluation and assessment frameworks. These involve more responsibility given to the school itself, through greater emphasis on school self-evaluation; greater importance of external school evaluation as accountability requirements increase; more emphasis on school leadership and its appraisal as the pedagogical role of school leaders is consolidated; the emergence of formal systems of teacher appraisal; the expansion of student standardised assessment to monitor learning outcomes; the growing importance of performance data, particularly relating to student outcomes, to inform school and classroom practices as well as system-level policies; and the growing emphasis on the use of data for formative assessment.

Educational measurement and indicators development are rising in importance

The introduction of national standardised assessments for students in a large number of countries reflects the stronger focus on measuring student outcomes. These make data on student learning outcomes available, providing a picture of the extent to which student learning objectives are being achieved, and they grant the opportunity to compare student learning outcomes across individual schools, regions of the country and over time. Also, for the purpose of monitoring education systems and evaluating school performance, data are increasingly complemented by a wide range of education indicators based on demographic, administrative and contextual data collected from individual schools. Most countries have developed comprehensive national indicator frameworks. It is now common practice to report statistics and indicators in education in an annual publication. International benchmarking is also increasingly common.

Larger and more varied uses are given to evaluation and assessment results

Countries are giving a more varied use to evaluation and assessment results. There is a growing interest in using evaluation results for formative purposes. School leaders, teachers and policy makers are more and more using evaluation results to identify areas where schools are performing well, and where they may need to improve. These data may help shape policy and/or school management decisions on resource distribution,

curriculum development and definition of standards, or strategies for professional development. Another increasingly marked focus is the use of evaluation and assessment results to hold policy makers, school leaders and teachers accountable.

Accountability as a purpose of evaluation and assessment is gaining in importance

Countries are increasingly using evaluation and assessment for accountability purposes. This can take a variety of forms. First, there is a growing trend of public reporting, including the publication of standardised student assessment results at the school level for use by parents, government officials, the media and other stakeholders, the publication of school inspection reports, school annual reports, and system level reports providing an assessment of the state of education. Second, evaluation and assessment results are increasingly used to reward or sanction the performance of individual school agents. This goes alongside the expansion of school external evaluation and teacher appraisal procedures. A number of countries have instituted systems whereby either schools, school leaders or teachers receive rewards for their good performance or are the subject of sanctions for underperformance.

There is greater reliance on educational standards

The focus on student learning outcomes has, in many countries, driven the establishment or underlined the importance of educational standards for the quality of the work of schools and school agents, and encouraged means for monitoring progress towards those standards. Educational standards refer to descriptions of what students should know (content standards) and be able to do (performance standards) at different stages of the learning process. In many countries, there is growing emphasis on the development and use of ambitious educational standards as the basis of assessment and accountability. By creating a set of standards against which student performance can be measured, countries aim to assess students against a desired measurable outcome.

Assessment is becoming more international

National education debates are increasingly shaped by international comparisons, particularly of student performance in international student surveys. The growing availability of internationally comparable data on student performance has, in important ways, influenced national discussions about education and fostered education policy reforms in countries. International comparative data put countries under pressure to attain higher levels of performance building on policies identified as potentially effective in high-performing countries. The expansion of international assessment has also significantly contributed for some countries to introduce national standardised assessments.

Assessment involves greater technological sophistication

The expansion of assessment, particularly the spreading out of standardised student assessment, as well as the management of the data it generates has greatly benefited from greater capacity of information and communication technologies. Improvements include more individualised assessment approaches, better assessment of cognitive skills such as problem solving, capacity for rapidly marking large-scale assessments, reliability in marking and reduced cost to administer student assessment. Other examples include the development of rapid-assessment – a computer-facilitated approach to frequent, brief

formative student assessment, more sophisticated value-added models to determine a school's or a teacher's contribution to student learning, and data information systems providing new opportunities for information sharing across school agents.

Main policy challenges

In all countries, there is widespread recognition that evaluation and assessment frameworks are key to building stronger and fairer school systems. Countries also emphasise the importance of seeing evaluation and assessment not as ends in themselves, but instead as important tools for achieving improved student outcomes. However, there are a range of challenges in ensuring that evaluation and assessment reach such ultimate objective. Although each country context is unique, some common policy challenges emerge. These are listed in Table 1.

Table 1. Main challenges in evaluation and assessment

Domain	Main challenges
The evaluation and assessment framework	Building a coherent and integrated evaluation and assessment framework Balancing the accountability and development functions of evaluation and assessment Ensuring articulations within the evaluation and assessment framework Securing links with classroom practice Finding a desirable measure of national consistency as against local diversity Developing competencies for evaluation and assessment and for using feedback Overcoming the challenge of implementation
Student assessment	Aligning educational standards and student assessment Finding a balance between summative and formative assessment Balancing external assessments and teacher-based assessments in the assessment of learning Developing fair assessments to all student groups Designing large-scale assessments that are instructionally useful Ensuring fairness in assessment and marking across schools Securing informative reporting of student assessment results
Teacher appraisal	Developing a shared understanding of high-quality teaching Balancing the developmental and accountability functions of teacher appraisal Accounting for student results in the appraisal of teachers Developing adequate skills for teacher appraisal Using teacher appraisal results to shape incentives for teachers
School evaluation	Aligning the external evaluation of schools with school self-evaluation Ensuring the centrality of the quality of teaching and learning Balancing information to parents with fair and reasonable public reporting on schools Building competence in the techniques of self-evaluation and external school evaluation Improving the data handling skills of school agents
The appraisal of school leaders	Developing school leadership appraisal as an integral part of the evaluation and assessment framework Developing a clear understanding of effective school leadership Placing pedagogical/learning-centred leadership at the heart of school leadership appraisal Combining the improvement and accountability functions of school leadership appraisal Ensuring that all school leaders have opportunities for professional feedback Using appraisal results to shape incentives for school leaders
Education system evaluation	Meeting information needs at the system level Monitoring key student learning outcomes Securing comparability over time and across schools Developing analytical capacity to use education system evaluation results for improvement Communicating education system evaluation results clearly and comprehensively Maximising use of system-level information

Main policy directions

To meet the challenges outlined above, a number of policy options are suggested across the areas analysed – the evaluation and assessment framework, student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation. Table 2 summarises the main policy directions (see Annex B for a complete list of policy directions). Not all of the policy directions apply equally to all countries. In a number of cases many, or most, of the policy suggestions are already in place, while for other countries they may have less relevance because of different social, economic and educational structures and traditions. This is a challenging agenda, but tackling one area without appropriate policy attention to inter-related aspects will lead to only partial results. Nevertheless, it is difficult to address all areas simultaneously, and resource constraints mean that trade-offs are inevitable.

Table 2. Main policy directions

Policy objective	Main policy directions
The evaluation and assessment framework **Embracing a holistic approach**	Integrate the evaluation and assessment framework Align the evaluation and assessment framework with educational goals and student learning objectives Secure links to the classroom and draw on teacher professionalism Promote national consistency while giving room for local diversity Establish articulations between components of the evaluation and assessment framework Place the students at the centre of the evaluation and assessment framework Sustain efforts to improve capacity for evaluation and assessment Engage stakeholders and practitioners in the design and implementation of evaluation and assessment policies
Student assessment **Putting the learner at the centre**	Ensure a good balance between formative and summative assessment Establish safeguards against an overreliance on standardised assessments Draw on a variety of assessment types to obtain a rounded picture of student learning Support effective formative assessment processes Ensure the consistency of assessment and marking across schools Ensure that student assessment is inclusive and responsive to different learner needs Put the learner at the centre and build students' capacity to engage in their own assessment Maintain the centrality of teacher-based assessment and promote teacher professionalism Engage parents in education through adequate reporting and communication
Teacher appraisal **Enhancing teacher professionalism**	Resolve tensions between the developmental and accountability functions of teacher appraisal Consolidate regular developmental appraisal at the school level Establish periodic career-progression appraisal involving external evaluators Establish teaching standards to guide teacher appraisal and professional development Prepare teachers for appraisal processes and strengthen the capacity of school leaders for teacher appraisal Ensure that teacher appraisal feeds into professional development and school development Establish links between teacher appraisal and career advancement decisions
School evaluation **From compliancy to quality**	Ensure the focus for school evaluation is the improvement of teaching, learning and student outcomes Evaluate and adapt external school evaluation to reflect the maturity of the school evaluation culture Raise the profile of school self-evaluation and align external school evaluation with school self-evaluation Develop nationally agreed criteria for school quality to guide school evaluation Strengthen school principals' capacity to stimulate an effective school self-evaluation culture Promote the engagement of all school staff and students in school self-evaluation Promote the wider use of the results of external school evaluation Report a broad set of school performance measures with adequate contextual information

Table 2. Main policy directions (*continued*)

Policy objective	Main policy directions
The appraisal of school leaders **Fostering pedagogical leadership in schools**	Promote the effective appraisal of school leaders within the broader assessment and evaluation framework Develop a common leadership framework or set of professional standards for school leaders Promote the appraisal of pedagogical leadership together with scope for local adaptation Build capacity for effective school leader appraisal Ensure school leader appraisal informs professional development Consider career advancement opportunities to reward successful school leaders
Education system evaluation **Informing policies for system improvement**	Ensure a broad concept of education system evaluation within the evaluation and assessment framework Ensure policy making is informed by high-quality measures, but not driven by their availability Develop a national education indicator framework and design a strategy to monitor student learning standards Ensure the collection of qualitative information on the education system Ensure collection of adequate contextual information to effectively monitor equity Establish and secure capacity for education system evaluation Strengthen analysis of education system evaluation results for planning and policy development

Common policy themes

Despite the major differences and traditions across countries, they share some common policy priorities.

Fostering synergies within the evaluation and assessment framework

The full potential of evaluation and assessment will not be realised until the framework is fully integrated and is perceived as a coherent whole. This requires a holistic approach to building a complete evaluation and assessment framework in view of generating synergies between its components, avoiding duplication of procedures and preventing inconsistency of objectives.

At the outset, it might prove useful to develop a strategy or framework document that conceptualises a complete evaluation and assessment framework and articulates ways to achieve the coherence between its different components. The strategy should establish a clear rationale for evaluation and assessment and a compelling narrative about how evaluation and assessment align with the different elements in the education reform programme. It should describe how each component of the evaluation and assessment framework can produce results that are useful for classroom practice and school development activities. The strategy could also contribute to clarifying responsibilities of different actors for the different components and allow for better networking and connections between the people working on evaluation and assessment activities. As such, it should also create the conditions for a better articulation between the different levels of educational governance, including evaluation agencies and local education authorities.

Furthermore, the process of developing an effective evaluation and assessment framework should give due attention to: achieving proper articulation between the different evaluation components (e.g. school evaluation and teacher appraisal); warranting the several elements within an evaluation component are sufficiently linked (e.g. teaching standards and teacher appraisal; external school evaluation and school self-evaluation); and ensuring processes are in place to guarantee the consistent application of

evaluation and assessment procedures (e.g. consistency of teachers' marks). This is in addition to proper links to initial teacher education and strategies for professional development; situating education system evaluation in the broader context of public sector performance requirements; and ensuring references for evaluation and assessment are well aligned with student learning objectives.

Aligning student learning goals with evaluation and assessment

A critical aspect in the effectiveness of the evaluation and assessment framework is its proper alignment with educational goals and student learning objectives. This involves a range of aspects. First, it requires evaluation and assessment procedures to align with the main principles embedded in educational goals and student learning objectives. For instance, if educational goals are based on principles such as student-centred learning, collaborative work, achievement of competencies and assessment for learning then there should be greater emphasis on the developmental function of evaluation and assessment, involving more attention to student formative assessment, greater emphasis on self-reflection for all the school agents, greater focus on continuous improvement in teacher appraisal, and better use of results for feedback.

Second, evaluation and assessment procedures require direct alignment with student learning objectives. This implies designing fit-for-purpose student assessments which focus on the competencies promoted in student learning objectives, ensuring the overall evaluation and assessment framework captures the whole range of student learning objectives, and developing teaching and school management standards which are aligned with student learning objectives. Third, it is essential that all school agents have a clear understanding of education goals. This requires goals to be clearly articulated; the development of clear learning expectations and criteria to assess achievement of learning objectives; room for schools to exercise some autonomy in adapting learning objectives to their local needs; and collaboration among teachers and schools to ensure moderation processes which enhance the consistency with which learning goals are achieved. Fourth, it is essential to evaluate the impact of evaluation and assessment against student learning objectives on the quality of the teaching and learning. Particular attention should be given to identifying unintended effects as evaluation and assessment activities have considerable potential to determine the behaviour of school agents.

Focussing on the improvement of classroom practices and building on teacher professionalism

To optimise the potential of evaluation and assessment to improve what is at the heart of education – student learning – policy makers should promote the regular use of evaluation and assessment results for improvements in the classroom. All types of evaluation and assessment should have educational value, and be meaningful to those who participate in the evaluation or assessment. To this end, it is important that all those involved in evaluation and assessment at the central, local and school level have a broad vision of evaluation and assessment and of the need to bring together results from different types of evaluation and assessment activities to form rounded judgements about student learning, performance of school agents and practices within the school system and use evaluation and assessment information for further improvement.

This calls for an articulation of ways for the evaluation and assessment framework to generate improvements in classroom practice through the assessment and evaluation procedures which are closer to the place of learning. Evaluation and assessment have no

value if they do not lead to the improvement of classroom practice and student learning. An important step in this direction is a national reflection about the nature and purpose of evaluation components such as school evaluation, school leader appraisal, teacher appraisal and student formative assessment within the overall education reform strategy and the best approaches for these evaluation components to improve classroom practices. Impacting classroom practice is likely to require the evaluation and assessment framework to place considerable emphasis on its developmental function. Channels which are likely to reinforce links to classroom practice include: an emphasis on teacher appraisal for the continuous improvement of teaching practices; ensuring teaching standards are aligned with student learning objectives; involving teachers in school evaluation, in particular through conceiving school self-evaluation as a collective process with responsibilities for teachers; focussing school evaluation on the quality of teaching and learning and their relationship to student learning experiences and outcomes; promoting the appraisal of the pedagogical leadership of school leaders; ensuring that teachers are seen as the main experts not only in instructing but also in assessing their students, so teachers feel the ownership of student assessment and accept it as an integral part of teaching and learning; building teacher capacity for student formative assessment; and building teachers' ability to assess against educational standards.

The central agent in securing links between the evaluation and assessment framework and the classroom is the teacher. This highlights the importance for evaluation and assessment frameworks to draw on the professionalism of teachers in ensuring evaluation and assessment activities result in authentic improvement of classroom practices and student learning. In addition, establishing links between evaluation and assessment and classroom learning requires establishing clear roles for local structures – school management, school supervision, local education authorities – in the implementation of evaluation and assessment policies. The point is that the fulfilment of the developmental function of evaluation and assessment requires articulation at the local level.

Effectively conceiving the accountability uses of evaluation and assessment results

Evaluation and assessment provide a basis for monitoring how effectively education is being delivered to students and for assessing the performance of systems, schools, school leaders, teachers and students, among others. They can serve as an instrument for the accountability of school agents when the results of an evaluation or assessment have stakes for school agents such as linkages to career advancement or salary progression, one-off rewards, sanctions, or simply information to parents in systems based on parental school choice. By measuring student outcomes and holding teachers, school leaders and schools responsible for results, accountability systems intend to create incentives for improved performance and identify "underperforming" schools and school agents.

At the same time, high-stakes uses of evaluation and assessment results might lead to distortions in the education process as a result of school agents concentrating on the measures used to hold them accountable. For instance, if those measures are based on student standardised tests, this might include excessive focus on teaching students the specific skills that are tested, narrowing the curriculum, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and potentially even manipulation of results. Also, when the framework tends to stress the accountability function there is a risk that evaluation and assessment are perceived mostly as instruments to hold school agents accountable, to

"control" and assess compliance with regulations. An additional challenge is that the developmental function of evaluation and assessment might be hindered in processes which stress the accountability function as the high stakes involved will limit the ability of school agents to openly reveal their weaknesses in view of receiving feedback to improve their practices.

As a result, it is important to design the accountability uses of evaluation and assessment results in such a way these undesired effects are minimised. This involves safeguards against excessive emphasis on particular measures, such as student standardised tests, to hold school agents accountable and drawing on a broad range of assessment information to make judgements about performance; communicating that the ultimate objective of evaluation and assessment is to enhance student outcomes through the improvement of practices at the different levels of the school system; building on a variety of evaluation and assessment procedures achieving each a well-identified distinct function; ensuring that the publication of quantitative data is perceived as fair by schools and set in a wider set of evidence; and conceiving individual performance-based rewards for school personnel as career advancement opportunities and non-monetary rewards.

Placing the student at the centre

Given that the fundamental purpose of evaluation and assessment is to improve the learning of the students, a key principle is to place the students at the centre of the framework. This translates into teaching, learning and assessment approaches which focus on students' authentic learning. Students should be fully engaged with their learning, contributing to the planning and organisation of lessons, having learning expectations communicated to them, assessing their learning and that of their peers, and benefitting from individualised support and differentiated learning. To become lifelong learners, students need to be able to assess their own progress, make adjustments to their understandings and take control of their own learning. Student feedback to teachers can also be used for teacher formative appraisal. In addition, it is important to build community and parental involvement and an acceptance of learning and teaching as a shared responsibility. A particularly important priority for some countries is to reduce the high rates of grade repetition. There are alternative ways of supporting those with learning difficulties in the classroom.

In addition, evaluation and assessment should focus on improving student outcomes and achieving student learning objectives. This should be reflected in the priorities for national monitoring, the importance of evidence on student performance for school evaluation and teacher appraisal, the value of clear reporting on student results, and the emphasis on feedback for improving student learning strategies. There is also the increasing recognition that the monitoring of student outcomes must extend beyond knowledge skills in key subject areas and include broader learning outcomes, including students' critical thinking skills, social competencies, engagement with learning and overall well-being.

Going beyond measurement in educational evaluation

As described earlier, measures of student learning are becoming increasingly available (in particular through national standardised assessments) and most countries have developed education indicator frameworks. Performance in schools is increasingly judged on the basis of effective student learning outcomes. This is part of the general shift

to outcome measures in the public sector. The advantage is that student outcomes become the focal point for analysis.

An imperative is that measures of performance are broad enough to capture the whole range of student learning objectives. However, it is not always possible to devise indicators and measures of good quality across all the objectives of the education system. Hence, it needs to be recognised that policy making at the system level needs to be informed by high-quality data and evidence, but not driven by the availability of such information. Qualitative studies as well as secondary analysis of the available measures and indicators are essential information to take into account in policy development and implementation. Qualitative approaches include the narrative provided by external school evaluation reports, key stakeholder feedback on broader outcomes (e.g. school climate, student engagement), and qualitative appraisal of teachers and school leaders. The qualitative aspects can feed into the policy debate by providing evidence on a broader set of student learning outcomes, as well as help shed light on some of the factors associated with student learning outcomes.

Building capacity for evaluation and assessment

The development of an effective evaluation and assessment framework involves considerable investment in developing competencies and skills for evaluation and assessment at all levels. Hence, an area of policy priority is sustaining efforts to improve the capacity for evaluation and assessment. Depending on country specific circumstances, areas of priority might be: developing teachers' capacity to assess against student learning objectives; improving the skills of teachers for formative assessment; improving the data handling skills of school agents; or developing expertise for teacher appraisal and school evaluation, including ensuring that designated evaluators are qualified for their role. Capacity building through adequate provision of initial teacher education and professional development should be a priority making sure provision is well aligned with the national education agenda. This should go alongside the development of training and competency descriptions for key people within the evaluation and assessment framework.

There is also a need to reinforce the pedagogical leadership skills of school directors as their role in many countries still retains a more traditional focus on administrative tasks. The objective is that school leaders operate effective feedback, coaching and appraisal arrangements for their staff and effectively lead whole-school evaluation processes. Peer learning among schools should also be promoted. In addition, there needs to be strong capability at the national level to steer evaluation and assessment. This can be ensured through the establishment of agencies with high levels of expertise which have the capacity to foster the development of skills for evaluation and assessment across the system. Such agencies could provide important leadership in modelling and disseminating good practice within the evaluation and assessment framework.

A further strategy involves initiatives at the central level to build up a knowledge base, tools and guidelines to assist evaluation and assessment activities. These typically include detailed plans to implement student learning objectives, including guidelines for schools and teachers to develop student assessment criteria. Other examples are tools for teachers to use in the assessment of their students (e.g. test items banks), Internet platforms proposing formative teaching and learning strategies, tools for the self-appraisal of teachers, instruments for school leaders to undertake teacher appraisal, and resources for school self-evaluation.

Designing evaluation and assessment procedures which are fit for purpose

Establishing clarity about the purposes and appropriate uses of different evaluations and assessments is important to ensure that evaluation and assessment frameworks optimally contribute to improvements at the classroom, school and system level. A key goal for countries is to develop, for each component of the evaluation and assessment framework, a clear vision and strategy where different approaches developed nationally and locally each serve a clearly defined purpose and the format of the evaluation or assessment is aligned to that particular purpose. For evaluation or assessment to be meaningful, it must be well-aligned to the type of skills and competencies that are valued. Coherent evaluation and assessment frameworks should aim to align student learning objectives, practices in the school system, and evaluation and assessment around key learning goals, and include a range of different evaluation and assessment approaches and formats, along with opportunities for capacity building at all levels.

More specifically, because standardised central student assessment is a relatively new phenomenon in many OECD countries, it is important to be clear about its purposes, to develop large-scale assessments over time to be able to accommodate the purposes that are reasonable, point out inappropriate uses and provide guidance for the way in which these assessments can be used as part of a broader assessment framework. Also, to build a systematic and coherent system of teacher appraisal, it is important that the aspects it seeks to monitor and improve are clear, that the approaches to appraisal are adapted to the different stages of a teachers' career and in line with the purposes they are aiming to achieve. Similarly, the fundamental purpose of school evaluation needs to be clearly and consistently understood across the school system. For instance, external school evaluation can be part of the strategy to bring about general improvement across all schools or, more narrowly, it can focus on "underperforming schools".

Evaluation and assessment systems also need to underline the importance of responding to individual needs and school community contexts, and design evaluation and assessment strategies that suit the needs of different learner groups or distinct schools agents.

Balancing national consistency with meeting local needs

In order to contribute to national reform agendas, a certain degree of national consistency of approaches to evaluation and assessment is desirable. This is likely to provide greater guarantees that evaluation and assessment practices are aligned with national student learning objectives. However, in certain countries, there are strong traditions of local ownership – at the jurisdiction level (federal systems), local level (region or municipality), or school level. In these cases, a high degree of autonomy is granted in school policies, curriculum development and evaluation and assessment. There is an understanding that shared or autonomous decision making and buy-in from those concerned are essential for the successful implementation of evaluation and assessment policy. It is also clear that local actors are in a better position to adapt evaluation and assessment policies to local needs.

Hence, the evaluation and assessment framework will need to find the right balance between national consistency and local diversity. A possible approach is to agree general principles for the operation of procedures such as school evaluation, teacher appraisal, school leader appraisal and student assessment while allowing flexibility of approach within the agreed parameters to better meet local needs. The principles agreed should come along with clear goals, a range of tools and guidelines for implementation. In

decentralised systems, it is also important to encourage the different actors to co-operate, share and spread good practice and thereby facilitate system learning, development and improvement.

Implementing evaluation and assessment policy successfully

The process of evaluation and assessment policy design and implementation involves a number of challenges to yield sound results. Ideally, evaluation and assessment policy would need to be based upon informed policy diagnosis, drawn on best practice, backed up by adequate research evidence, and consistent – both intrinsically and with other education policies. Of equal importance is consensus-building among the various stakeholders involved – or with an interest – in educational evaluation. This should go alongside the involvement of practitioners such as school leaders and teachers in the design, management and analysis of evaluation and assessment policies.

In order to build consensus, it is important that all stakeholders see proposed evaluation and assessment policies within the broader policy framework and strategy. Indeed, individuals and groups are more likely to accept changes that are not necessarily in their own best interests if they understand the rationale for these changes and can see the role they should play within the broad evaluation and assessment framework. There is therefore much scope for government authorities to foster the chances of successful policy implementation, by improving communication on the long-term vision of what is to be accomplished for evaluation and assessment as the rationale for proposed reform packages.

Other approaches for successful policy implementation include the use of pilots and policy experimentation when needed, opportunities for education practitioners to express their views and concerns as evaluation and assessment policies are implemented, the communication of key evaluation and assessment results to stakeholders, developing expertise and capacity for evaluation and assessment across the system, reducing excessive bureaucratic demands on schools, and ensuring sufficient resources are provided for implementation.

Chapter 1

The focus on evaluation and assessment

The OECD undertook a major international study of evaluation and assessment policies in school systems: the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Drawing on the experiences of 25 countries around the world, and extensive data and research, the OECD Review analysed the key factors involved in effective evaluation and assessment policies, and developed policy options for countries to consider.

Over at least the last 30 years, evaluation and assessment have become an increasingly significant feature of the educational landscape in countries across the world. Their nature and purpose remain varied, reflecting national traditions, infrastructure and practices, broader educational policy and political agreements. This chapter discusses why evaluation and assessment policies are high on national agendas, describes the methodology used in the Review, and outlines the organisation of this report.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The growing importance of evaluation and assessment

There is widespread recognition that evaluation and assessment arrangements are key to both improvement and accountability in school systems. This is reflected in their increasing importance in national education agendas (see Chapter 2). As countries strive to transform their educational systems to prepare all young people with the knowledge and skills needed to function in rapidly changing societies, some common policy trends can be observed in one form or another in most OECD countries, including decentralisation, school autonomy, greater accountability for outcomes and a greater knowledge management capacity. Decentralisation and school autonomy are creating a greater need for the evaluation of schools and school agents while greater IT capacity allows for the development and analysis of large-scale student assessments as well as individualised assessment approaches. Results from evaluation and assessment are becoming critical to knowing whether the school system is delivering good performance and to providing feedback for further development. Evaluation and assessment are instrumental in defining strategies for improving school practices with the ultimate goal of enhancing student outcomes. These developments are having a strong influence in the way in which policy makers monitor system, school, school leader, teacher and student performance.

Countries increasingly use a range of approaches to the evaluation and assessment of students, teachers, school leaders, schools and education systems. These are used as tools for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school, school leadership and teaching practices. Strong emphasis is being placed on better equipping and encouraging teachers to carry out self-appraisal and student formative assessment, on providing the incentives and means for school self-evaluation, on encouraging "value-added" evaluation and on more regular standardised testing of students and national monitoring of the overall system. However, countries often face difficulties in implementing evaluation and assessment procedures. These may arise as a result of poor policy design, lack of analysis of unintended consequences, little capacity for educational agents to put procedures into practice, lack of an evaluation culture, or deficient use of evaluation results.

Against this background, the OECD Education Policy Committee launched the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes in late 2009 to provide analysis and policy advice to countries on how evaluation and assessment arrangements can be embedded within a consistent framework to bring about real gains in performance across the school system.

Methodology

This report is concerned with evaluation and assessment policies in school systems that can help countries achieve their educational goals and student learning objectives. It draws on a major study, the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes, conducted in collaboration with OECD countries and under the guidance of the OECD's Group of National Experts on Evaluation and Assessment.

The Review was based on volunteer countries working collaboratively with each other and with the OECD Secretariat. It was designed to respond to the strong interest in evaluation and assessment issues evident at national and international levels. The Review looked at the various components of evaluation and assessment frameworks that countries

use with the objective of improving student outcomes. These include student assessment, teacher appraisal, school evaluation, school leader appraisal and system evaluation. The Review focused on primary and secondary education.[1] It involved examining country-specific issues and policy responses in strengthening the contribution of evaluation and assessment frameworks to the improvement of student outcomes to generate insights and findings relevant to OECD countries as a whole.[2]

The overall purpose of the Review was to explore how evaluation and assessment frameworks can be used to improve the quality, equity and efficiency of school education. The overarching policy question was "How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools?" The Review further concentrated on five key issues for analysis: *(i)* designing a systemic framework for evaluation and assessment; *(ii)* ensuring the effectiveness of evaluation and assessment procedures; *(iii)* developing competencies for evaluation and for using feedback; *(iv)* making the best use of evaluation results; and *(v)* implementing evaluation and assessment policies.

The project was overseen by the Group of National Experts on Evaluation and Assessment, which was established as a subsidiary body of the OECD Education Policy Committee in order to guide the methods, timing and principles of the Review.

Twenty-five countries were actively engaged in the Review. The fact that so many countries took part indicates that evaluation and assessment issues are a priority for public policy, and likely to become even more so in future years. Participating countries covered a wide range of economic and social contexts, and among them they illustrated quite different approaches to evaluation and assessment in school systems. This allowed a comparative perspective on key policy issues. These countries prepared a detailed background report, following a standard set of guidelines. Fifteen countries also opted for a detailed Review, undertaken by a team consisting of members of the OECD Secretariat and external experts. This resulted in the preparation of 15 Country Review Reports containing an analysis of evaluation and assessment policies and policy recommendations in the concerned countries.[3] The countries actively engaged in the Review were:[4]

- Preparation of a Country Background Report (25 countries, involving 26 reports): Australia, Austria, Belgium (Flemish Community), Belgium (French Community), Canada,[5] Chile, the Czech Republic, Denmark, Finland, France, Hungary, Iceland, Ireland, Italy, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Sweden and the United Kingdom (Northern Ireland).
- Country Review countries (15 countries): Australia, Belgium (Flemish Community), Chile, the Czech Republic, Denmark, Italy, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Portugal, the Slovak Republic, Sweden and the United Kingdom (Northern Ireland).

There are some striking differences among countries in regard to their evaluation and assessment frameworks, as illustrated by:

- Student assessment: In primary education, students are not awarded marks in Denmark, Iceland, New Zealand, Norway and Sweden whereas Hungary, Italy, Mexico, the Netherlands, Poland and the Slovak Republic rely primarily on numerical marks for formal reporting.

- Teacher appraisal: In Australia, Chile, Korea, Portugal and the United Kingdom, teachers undergo formal processes of appraisal as part of their performance management whereas in Denmark, Finland, Iceland, Norway and Sweden feedback on teacher performance is provided in more informal settings at the school level.
- School evaluation: While in systems such as Australia, the Flemish Community of Belgium, the Czech Republic, the Netherlands, New Zealand, Portugal, the Slovak Republic, Sweden and the United Kingdom there is a tradition of external school evaluation (inspections or reviews), in countries such as Chile, Mexico and the United States school evaluation is mostly based on comparable performance measures.
- School leader appraisal: In Austria, Estonia, Iceland, Ireland, Italy and Luxembourg no requirements exist for school leader appraisal in contrast to the mandatory periodic appraisal systems in countries such as France, Israel, Korea, New Zealand, Portugal and Slovenia.
- System evaluation: While in systems such as Australia, Austria, the Flemish Community of Belgium, Chile, Iceland, New Zealand and Norway the use of stakeholder surveys for system evaluation is common, other systems such as the Czech Republic, France, Korea, Luxembourg and Mexico do not rely on them.

By documenting such differences among countries, and trying to understand their causes and consequences, comparative analysis can help to raise questions about long-established practices and help accumulate evidence on the impact of different policy approaches. Annex A details the processes involved in the OECD Review, the country reports and other documents produced, and the large number of organisations and people who contributed to the Review and to the preparation of this report.

Organisation of the report

This report is intended to add value to the wide range of materials produced through the Review (listed in Annex A) by drawing out its key findings and policy messages. This report seeks to:

- provide an international comparative analysis of evaluation and assessment policies in school systems
- provide a stock-take of current policies and practices in countries
- draw attention to effective and innovative policy initiatives in countries
- develop a comprehensive framework to guide the development of evaluation and assessment policies
- propose evidence-based policy options for the development of evaluation and assessment policies.

The contexts within which evaluation and assessment policy making operates can vary markedly across countries depending upon their historical traditions, educational cultures and economic conditions. Policy initiatives that work well in one national context are not necessarily transferable. The Review has attempted to be sensitive to this through an approach that analyses evaluation and assessment policies in relation to the values,

vision and organisation of school systems in different countries as well as the broader economic, social and political contexts in which they operate.

The report has seven further chapters. Chapter 2 provides an overview of context and trends in evaluation and assessment. Chapters 3-8 are concerned with the key substantive issues driving the project: the evaluation and assessment framework, including policy implementation (Chapter 3); student assessment (Chapter 4); teacher appraisal (Chapter 5); school evaluation (Chapter 6); the appraisal of school leaders (Chapter 7); and education system evaluation (Chapter 8). Each of these chapters discusses country practices, the main factors involved and their impact, and examples of innovative policy responses, as well as identifying policy options for countries to consider. Annex A details the process by which the Review was conducted, and the range of outputs in addition to this report. Annex B lists the main policy recommendations.

The report provides analysis on:

- how to develop a coherent evaluation and assessment framework within which different components serve their purpose well and together successfully combine development and accountability to improve student outcomes
- how to strengthen the use of evaluation and assessment results to enhance school and classroom practices so student outcomes are improved
- how to facilitate the implementation of evaluation and assessment policies.

It provides a description of design, implementation and use of evaluation and assessment procedures in countries, analyses strengths and weaknesses of different approaches, and offers recommendations for improvement including how results should be incorporated into policy and practice. In doing so, it synthesises research-based evidence on the impact of evaluation and assessment strategies.

The following chapters provide many examples of country initiatives in evaluation and assessment. A number of particularly innovative and promising initiatives are highlighted in self-contained boxes that provide more detail on the reforms. Nevertheless, due to space constraints, it has not been possible to provide all of the necessary detail, and readers are encouraged to consult the relevant Country Background Reports, Country Review Reports and research studies. All the documents produced through the project are available from *www.oecd.org/edu/evaluationpolicy*. It should be noted that country-specific information given in this report with no associated source or reference is taken from Country Background Reports and Country Review Reports produced through the Review.

Notes

1. The scope of the Review did not include early childhood education and care, apprenticeships within vocational education and training, and adult education.
2. The project's purposes, analytical framework and methodology are detailed in the OECD (2009) document entitled "OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Design and Implementation Plan for the Review" available at *www.oecd.org/edu/evaluationpolicy*.
3. The Country Review Reports were released as the publication series *OECD Reviews of Evaluation and Assessment in Education*.
4. However, to the extent they are covered by the OECD Education Database and by the academic and policy literatures, OECD countries less actively engaged in the Review are still considered in the analysis and feature in some of the report's figures and tables.
5. The Country Background Report for Canada focuses on policies and practices of the Atlantic provinces: Newfoundland and Labrador, Nova Scotia, Prince Edward Island, and New Brunswick.

Chapter 2

Trends in evaluation and assessment

This chapter provides the context for analysing evaluation and assessment policy. First, it describes the main trends within educational evaluation. It is apparent that evaluation and assessment are increasingly being considered as levers of change guiding improvement, accountability, educational planning and policy development within school systems. Countries are developing more comprehensive evaluation and assessment frameworks, placing greater emphasis on educational measurement and indicators development, giving growing prominence to accountability uses of results, and relying increasingly on educational standards.

Second, the chapter reviews the contextual factors shaping the development of evaluation and assessment in school systems. Evaluation and assessment have gained in importance as a result of greater levels of school devolution, a stronger role for market-type mechanisms in education, the emergence of New Public Management, the growing imperative of an efficient use of public resources, the need to focus on "quality for all" and the rising importance of education in a global world. Other contextual factors influencing the development of evaluation and assessment frameworks include the rising expectations of the professionalism of teachers, more educated parents, the movement to advance the use of evidence-based decision making, technological advancements, the emergent commercial interests in education and the role of the media.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter provides the context for analysing evaluation and assessment policy. First, it describes the main trends within educational evaluation, with particular emphasis on the expansion and diversification of evaluation and assessment procedures. Second, it reviews the contextual factors shaping the development of evaluation and assessment in school systems. More specific trends within student assessment, teacher appraisal, school evaluation, school leader appraisal and system evaluation are analysed in the respective chapter.

Trends in evaluation and assessment

Expansion of educational evaluation in school systems

Increased prominence of evaluation and assessment in education policy

It is apparent that education policy is increasingly conferring a central strategic role to evaluation and assessment as indispensable tools for improvement, accountability, educational planning and policy development. In the last two decades, most countries have introduced a wide range of measures intended to improve evaluation and assessment at all levels from the student to the school system itself. These have done much to stimulate public awareness of evaluation and assessment and to develop an evaluation culture within school systems.

For example, Norwegian authorities set up a National Quality Assessment System (NKVS) for the education sector in 2004. NKVS provides access to a range of data intended to help schools, school owners and education authorities evaluate their performance and inform strategies for improvement. It includes national student assessments at key stages of education, a range of user surveys, a web-based School Portal, and a range of tools for schools' self-review (e.g. diagnostic "mapping tests"). With the establishment and development of NKVS, policy makers aimed to move policy attention away from inputs and processes to focus more on the outcomes of education (Nusche et al., 2011a). Similarly, Portugal has come far in developing the foundations of a framework for evaluation and assessment. National monitoring educational progress tests were launched in 2001 (then replaced by national examinations in 2011/12), a first cycle of external school evaluations was completed in the period 2006-11, a national system of teacher performance appraisal was launched in 2007 and the availability of national indicators on education has considerably expanded (Santiago et al., 2012a). In Mexico, as of the National Education Programme 2001-06, it was established that evaluation and assessment should be permanent and systematic, combining the involvement of internal and external agencies and be important management instruments to achieve improvement and accountability to society (Santiago et al., 2012b). In Australia, the 2008 National Education Agreement, which established a national framework for reform in education, reinforced the role of evaluation and assessment as key tools to achieve quality and equity in education, in particular with the introduction of the National Assessment Program – Literacy and Numeracy (NAPLAN) and the establishment of a set of reporting requirements for all schools (Santiago et al., 2011).

The expansion of educational evaluation results from increased demands for effectiveness, equity and quality in education so new economic and social needs are met. It is part of the effort across many countries to bring about rapid and sustained large-scale

educational reform with real gains in student outcomes (Campbell and Levin, 2009). Evaluation and assessment have also gained in importance as a result of pressures for governments to meet public accountability, ensure transparency within education systems and maintain public confidence in schooling.

Creation of dedicated agencies as part of new approaches to govern evaluation and assessment

In many OECD countries, the greater importance of evaluation and assessment in education policy has involved the creation of specifically dedicated agencies which assume a central role in the governance of the evaluation and assessment framework (see also Table 3.2 in Chapter 3). Denmark created a specific national authority to monitor compulsory education (the Quality and Supervision Agency, formerly the School Agency) in 2011 and an advisory body to evaluate priorities in compulsory education (The Council for Evaluation and Quality Development of Primary and Lower Secondary Education) in 2006 (Shewbridge et al., 2011a). In Mexico, a milestone in the development of evaluation and assessment was the creation of the National Institute for Educational Assessment and Evaluation (INEE) in 2002 by presidential decree as a public, decentralised agency to provide national guidance and direction in evaluation and assessment activities at the school level. It gained further autonomy in 2012 with the objective of reinforcing its technical leadership of evaluation and assessment and maintaining the independence of its judgement of the state of education in Mexico (Santiago et al., 2012b).

In Chile, the Quality of Education Agency was created in 2011 and started operating in 2012. It takes responsibility for evaluating the quality of learning provided by Chilean schools, including the evaluation of teachers, school leaders and school providers, in view of improving the quality and equality of education. The Agency evaluates individual schools against learning standards, makes information about the performance of individual schools publicly available, and supervises and supports schools with lower performance (Santiago et al., forthcoming). In the Flemish Community of Belgium, the Agency for Quality Assurance in Education and Training (AKOV) was established in 2009 to oversee all services related to quality improvement of education (Shewbridge et al., 2011b). In Italy, the creation in 2004 of the National Institute for the Evaluation of the Education System (INVALSI) led to the development of national student assessments since 2008. Similarly, in Ontario (Canada), the creation of the Education Quality and Accountability Office in 1996 was associated with the development of standardised student assessment to provide accountability and a gauge of quality in Ontario's publicly funded education system.

The creation of agencies dedicated to evaluation and assessment recognises the need for specialised expertise, the imperative of building adequate capacity to deliver evaluation and assessment policies and the necessity of introducing some independence *vis-à-vis* education authorities. Functions of evaluation and assessment agencies may include technical leadership (e.g. in developing evaluation instruments, guidelines, education indicators), implementation of evaluation and assessment procedures (e.g. national student assessments), the monitoring of the education system, the introduction of innovations on the basis of research results, the development of capacity for evaluation and assessment across the system, knowledge management (of results produced by evaluation and assessment activities) and the promotion of an evaluation culture.

Greater variety of evaluation and assessment activities

The expansion of educational evaluation was accompanied by considerable diversification of evaluation and assessment activities. Although educational evaluation within school systems is not a recent concern, it has traditionally focussed mostly on the assessment of students. As will be evident in subsequent chapters of this report, in recent years, countries are increasingly developing more comprehensive evaluation and assessment frameworks with more resources devoted to evaluation components other than student assessment.

For instance, within a context of growing levels of school autonomy, the responsibility for evaluation is increasingly given to the school itself, which involves greater emphasis of countries on school self-evaluation (see Chapter 6). At the same time, new accountability requirements for schools raise the importance of external school evaluation (see Chapter 6). In this context, some countries are now placing considerable emphasis on school leadership, developing reference standards for good leadership and establishing procedures to appraise school leaders (see Chapter 7). While less common in countries, there has also been a growing interest in developing formal teacher appraisal systems. Demands for instructional quality have led a number of countries to set up one form or another of teaching performance assessment (see Chapter 5). Some countries, such as the United States, are now using student standardised assessment results as an instrument to appraise individual teachers through the design of value-added models intended to measure the contribution of individual teachers to student learning (see Chapter 5). Also, as countries place greater emphasis on the monitoring of outcomes in public services, student standardised assessments play an increasingly important role in assessing learning outcomes in school systems (see below and Chapter 8). This is within a context in which countries demonstrate a growing interest in international benchmarks of student performance.

Another prominent development has been the growing importance of performance data, particularly relating to student outcomes, to inform school and classroom practices as well as system-level policies. This has generated a range of new practices in school systems related to the analysis of student outcomes. Strong emphasis is being placed on better equipping and encouraging teachers and other school agents to use data for formative assessment, on providing the incentives and means for student outcomes to be used in school self-evaluation, and on encouraging "value-added" approaches to assess the contribution of schools to student learning. Some countries have placed considerable emphasis on student information systems providing real-time access to student data on, for example, attendance, enrolment, marks and schedules. This is in a context in which education authorities have a growing concern of feeding back relevant information to school agents. The focus on student outcomes has also increasingly involved the establishment of longitudinal studies on the progression of individual students through the education system.

In the area of student assessment, in the last two decades there has been considerable policy attention to the consolidation of assessment for learning in the classroom and a growing support for the concept of *assessment as learning*, which focuses on students reflecting on and monitoring their own progress to inform future learning. In addition, technological advances have permitted student assessment to become more sophisticated, as is the case with computer-based adaptive assessment (see below).

The rise of educational measurement and indicators development

Student outcomes as the focal point for analysis

A major benefit of the stronger emphasis on evaluation and assessment has been the greater focus on improving student outcomes and achieving student learning objectives. This is reflected in the growing importance of student outcomes for system evaluation (increasingly relying on results of standardised student assessment and the international assessment of students), school evaluation (with school accountability increasingly tied to student outcomes) and teacher appraisal (with the exploration of direct links to student progress); the requirements for reporting publicly on student results; and the establishment of education national targets for student achievement including for particular groups of students. Performance in schools is increasingly judged on the basis of effective student learning outcomes. This is part of the general shift to outcome measures in the public sector. There is a greater emphasis on the use of student achievement data both to understand the balance between school, student and contextual data and to look at the school processes that appear to support improved achievements (Campbell and Levin, 2009).

The growing emphasis on measuring student outcomes

The introduction of national standardised assessments for students in a large number of countries reflects the stronger focus on measuring student outcomes. These make data on student learning outcomes available, providing a picture of the extent to which student learning objectives are being achieved, and they grant the opportunity to compare student learning outcomes across individual schools, regions of the country and over time. As put by Kellaghan and Greaney (2001), "The most remarkable development in assessment towards the end of the 20th century has probably been the growth in its use to measure the achievement outcomes of national systems of education, either considered uniquely (in national assessments) or in the context of the performance of other education systems (in international comparative studies of achievement)."

In Mexico, a ground-breaking development was the implementation of national standardised assessments: on the basis of a sample (EXCALE in 2005) and census-based (ENLACE in 2006). These made available data on student learning outcomes which, for the first time, provided a picture of the extent to which student learning objectives were being achieved. ENLACE also granted the opportunity to compare student learning outcomes across individual schools (Santiago et al., 2012b). As with Mexico, many OECD systems introduced, for the first time, central standardised assessments in core subjects in recent years, reflecting an impressive expansion of instruments to measure student outcomes. Examples include: Austria (2012), the Flemish Community of Belgium (2002), the French Community of Belgium (2009), Denmark (2009), Germany (2007), Hungary (2001), Iceland (2009), Ireland (2007), Israel (2002), Italy (2008), Japan (2007), Korea (2001), Luxembourg (2008), Norway (2004), Portugal (2001), Spain (2007) and the Slovak Republic (2004). In addition, the Czech Republic is piloting student standardised assessments in academic years 2011/12 and 2012/13 (see also Chapter 4; OECD, 2011; and Eurydice, 2009). Australia has also introduced standardised student assessment at the national level in 2008 with the National Assessment Program – Literacy and Numeracy (NAPLAN), even if centrally organised student assessments were in existence in several states and territories previously (Santiago et al., 2011). Similarly, Canada introduced standardised student assessment at the national level in 1996 through

the Student Achievement Indicators Program (SAIP), which was replaced by the Pan-Canadian Assessment Program as of 2007.

Other countries have had more extensive experience with national standardised student assessment. The United States organised the first National Assessment of Educational Progress (NAEP) in 1969. NAEP is a sample-based assessment whose results are designed to provide data on student achievement in various subjects and released as The Nation's Report Card. It may be considered the precursor to today's widespread use of student assessments as tools for holding educators accountable for student performance (Roeber, 1988, cited in Hamilton, 2003). The minimum competency testing movement of the 1970s emphasised the need to ensure that students demonstrated a grasp of basic skills and led to the first formal use of tests as tools to hold education administrators, students and teachers accountable for performance (Hamilton and Koretz, 2002). Over time there was a shift from the use of tests as measurement instruments designed to produce information to a reliance on tests to influence policy and instruction, a dual use that has continued to the present day (Hamilton, 2003). Other countries with an early experience with national student assessments include Ireland (national assessments in English reading and mathematics introduced to primary schools in 1972 and 1977 respectively), the Netherlands (standardised assessment in the majority of primary schools since 1970) and Sweden (formative national assessments in lower secondary education introduced in 1962) (Eurydice, 2009).

The proliferation of education indicators

For the purpose of monitoring education systems and evaluating school performance, data are increasingly complemented by a wide range of education indicators based on demographic, administrative and contextual data collected from individual schools. Datasets typically include information on students (type of enrolment, completion, absenteeism, age, gender, marks, socio-economic background), teachers (functions, qualifications, career status, age, gender, areas taught, teaching hours, absenteeism, remuneration), non-teaching staff (qualifications, age, gender, category), and schools (financial management, use of technology, organisation of learning). The emphasis is increasingly on output measures.

Most countries have developed comprehensive national indicator frameworks relying on data collection procedures at the school level. It is now also common practice to report statistics and indicators in education in an annual publication (e.g. *Statistical Yearbook of Education* and *Education in Figures*). In the Flemish Community of Belgium, for example, the Agency for Educational Services (AGODI, created in 2006) collects and analyses data on the state of education. In addition to the most conventional data collected at the school level (e.g. characteristics of teachers and students), it covers areas such as the operational resources and subsidies granted to schools, the re-structuring of schools and the labour market for teachers (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

International benchmarking is also increasingly common. A major driver for the collection of information on national education systems has been the joint international standardised data collection by UNESCO, OECD and EUROSTAT. It is also common to have countries publish education indicators in an international perspective. In Italy, the VALSIS project involves the analysis of international education system indicators to inform the Italian system evaluation framework. One output of the project is the creation of an electronic data bank with education system indicators.

Larger and more varied uses of evaluation and assessment results

Countries are giving a more varied use to evaluation and assessment results, including as a tool for understanding better how well students are learning, for providing information to parents and society at large about educational performance and for improving school and teaching practices. As will be analysed below, an increasingly marked focus is the use of evaluation and assessment results to hold policy makers, school leaders and teachers accountable. There is also a growing use of evaluation as a system steering tool. As put by Broadfoot and Black (2004), "In recent years the importance of assessment as a policy tool has grown enormously as governments have increasingly come to realise its powerful potential as a mechanism of state control." Evaluation procedures are now increasingly being considered as potential levers of change that can assist with decision making, resource allocation or school improvement.

There is a growing interest in using evaluation results for formative purposes. School leaders, teachers and policy makers are more and more using evaluation results to identify areas where schools are performing well, and where they may need to improve. These data may help shape policy and/or school management decisions on resource distribution, curriculum development and definition of standards, or strategies for professional development. School leaders and teachers can use evaluation data to change teaching, address ineffective programmes in their schools, and improve the functioning of the school in terms of increased student achievement (Schildkamp et al., 2012). There have also been profound changes in views on the role of classroom-based assessment. Assessment has traditionally been viewed as a tool for making summative judgements of student achievements. But increasingly, assessment is also seen as a tool for learning. Assessment, in this view, plays a "formative" role – allowing teachers to identify gaps in student learning and to adapt teaching appropriately (Looney, 2009).

The data generated by evaluation and assessment procedures is also increasingly motivating schools to engage in the corresponding analysis in view of improving student learning. Policy makers, school leaders and teachers are putting time and energy into making data accessible to schools and teachers and into using data, in training schools and teachers to use data for improvement, and in requiring schools and teachers to collect and publish data (Schildkamp et al., 2012). Greater reliance on self-evaluation by schools intensifies pressure for teachers and school leaders to possess the skills and know-how to collect and use information needed for diagnosis and performance measurement.

The growing prominence of accountability as a purpose of evaluation and assessment

Countries are increasingly using evaluation and assessment for accountability purposes. A central assumption in accountability is that substantial improvement necessitates that the school agents are held accountable for the outcomes they generate. By measuring student outcomes and holding teachers, schools and policy makers responsible for results, accountability systems intend to create incentives for improved performance and identify underperformance within school systems. Hargreaves and Shirley (2009), in their analysis of educational reforms, maintain that countries have gradually shifted from local and sampled assessments to high-stakes census testing for accountability purposes. Klenowski and Wyatt-Smith (2012) point to the political appeal of school-based accountability policies, in the sense of the clear need for politicians to be seen to deliver improved outcomes in education. Broadfoot and Black (2004) note, for example, that "decisions about assessment procedures – particularly those concerning

high-stakes testing of various kinds – are as often based on perceived political appeal as they are on a systematic knowledge on the scientific evidence concerning fitness for purpose" (as cited in Klenowski and Wyatt-Smith, 2012).

The school-based accountability movement emerged out of a desire, particularly seen in the United States and the United Kingdom beginning in the 1980s in the Reagan and Thatcher eras, to measure performance in the public and non-profit sectors (Figlio and Kenny, 2009). Elmore (2004) argues that test-based accountability has been more enduring in education than any other policy in the United States for at least the past 50 years and that it is unlikely to change in the near future. According to Hamilton (2003), the policy context in the United States is characterised by the use of tests in what may be called a test-based accountability system. These systems involve four major elements: *goals,* expressed in the form of standards; *measures* of performance (i.e. tests); *targets* for performance; and *consequences* attached to schools' success or failure at meeting the targets. As described by Figlio (2006), in the United States, the No Child Left Behind (NCLB) Act of 2001 solidified a national trend toward increased student testing for the purpose of evaluating public schools. Under NCLB, states must develop and administer rigorous curriculum-based standardised assessments to every student in a number of year levels. These tests must be used to evaluate schools, and in the case of schools receiving federal aid for disadvantaged students, aggregate performance on these assessments is associated with substantial rewards and sanctions. In 2009, the federal government of the United States launched the Race to the Top (RTT) initiative as a competition among states to access substantial federal funding. The RTT provides funding for states which implement a range of policies such as performance-based rewards for teachers and school leaders, adoption of common nationwide standards (from the Common Core State Standards initiative), and development of high-quality standardised student assessments.

As another example, in Ireland, there has been a drive for greater accountability in recent years. Examples of the move towards accountability include the introduction of regular whole-school inspection to secondary schools in 2003, the publication of school inspection reports in 2006, and the introduction of mandatory standardised testing in primary schools in 2007. The National Strategy for Literacy and Numeracy outlines additional accountability measures such as the development of national standards of students' achievement and the collection of national data on student achievement. The strategy also requires schools to provide parents with adequate, meaningful and clear assessment information on their child's progress (Irish Department of Education and Skills, 2012).

The accountability uses of evaluation and assessment can take a variety of forms. First, evaluation and assessment exist in an environment where there is a growing trend of public reporting (see also Chapter 6). This consists, for example, in publishing standardised student assessment results at the school level for use by parents, government officials, the media and other stakeholders. Not only does this serve the purpose of providing information on education system performance to the general public, but the results are often used by stakeholders to take action as with school parental choice. For example, in Chile national student assessment results (System for Measuring the Quality of Education, SIMCE) are published, inform the school voucher system and have contributed to placing education on the public agenda. In Australia, NAPLAN (National Assessment Program – Literacy and Numeracy) results are published on an individual school basis on the My School website, where the public can access performance and other data on schools across Australia. English schools' performance has also been

reported since 1988. The general move towards greater transparency of results includes the publication of school inspection reports, school annual reports and system level reports providing an assessment of the state of education.

Second, evaluation and assessment results are increasingly used to reward or sanction the performance of individual school agents. This goes alongside the expansion of school external evaluation and teacher appraisal procedures. A number of countries have now instituted systems whereby either schools or teachers receive rewards for their good performance (e.g. in the form of financial one-off packages, additional resources including opportunities for professional development, faster career advancement, opportunities for promotion), or are the subject of sanctions for underperformance (e.g. school shutdown, career stagnation, removal from post). The emergence of value-added techniques has also strengthened the interest in the use of student assessment results for evaluating and rewarding individual teachers and school leaders (Hout and Elliott, 2011).

Another development is the greater variety of accountabilities for school agents. For instance, in New Zealand, schools have multiple accountabilities – to their communities, the Ministry of Education, the Education Review Office, the New Zealand Teaching Council and the New Zealand Qualifications Authority (New Zealand Ministry of Education, 2010).

Greater reliance on educational standards

The focus on student learning outcomes has, in many countries, driven the establishment or underlined the importance of educational standards for the quality of the work of schools and school agents, and encouraged means for monitoring progress towards those standards. Educational standards refer to descriptions of what students should know (content standards) and be able to do (performance standards) at different stages of the learning process. In many countries, there is growing emphasis on the development and use of ambitious educational standards as the basis of assessment and accountability. By creating a set of standards against which student performance can be measured, countries aim to assess students against a desired measurable outcome. Examples of countries which implemented national educational standards are Australia, Austria, Belgium, Denmark, France, Luxembourg, Norway, the United Kingdom and the United States. By setting national or common standards, student outcomes can be more easily controlled for quality and they are more comparable (Wang et al., 2006). The movement towards comparing student outcomes to standards also has had a role in motivating countries to administer national standardised assessments.

At the same time, countries have adapted the key elements of standards-based systems to their own educational contexts and cultures – how they define standards, how they balance incentives and support, and how they measure school and student performance. Educational standards vary a good deal in specificity and emphasis across countries (see Chapter 4).

Internationalisation of assessment

National education debates are increasingly shaped by international comparisons, particularly of student performance in international student surveys (see also Chapter 8). These include student assessments conducted by the International Association for the Evaluation of Educational Achievement (IEA) (e.g. Progress in Reading Literacy Skills survey, PIRLS; Trends in Mathematics and Science Skills survey, TIMSS), the OECD's

Programme for International Student Assessment (PISA) or UNESCO's Latin American Laboratory for Assessment of the Quality of Education (LLECE). As explained by Bonnet (2004), "The conception of educational evaluation has changed over the years, moving from an essentially national to a more international perspective. It has also taken a new dimension with the fundamental role it now plays in Europe. In parallel with national developments countries started to show interest in comparative evaluation at the international level." Some education systems have been considerably shaken after publication of international comparative scores. For example, the first results of PISA published in December 2001 jolted Denmark and its education community. They provided evidence that one of the most expensive education systems in the world was performing at a level that, when compared to the outcomes observed in other OECD countries, was only average. The PISA results were at odds with the widely shared but poorly substantiated belief that Danish schools were the best in the world (OECD, 2008).

The growing availability of internationally comparable data on student performance has, in important ways, influenced national discussions about education and fostered education policy reforms in countries. International comparative data put countries under pressure to attain higher levels of performance building on policies identified as potentially effective in high-performing countries. A wide range of education reforms are triggered in OECD countries by student results in international assessments. For instance, in France, an application decree links the 2005 introduction of the common core competencies to the results of French 15-year-olds in PISA (Dos Santos and Rakocevic, 2012). Some countries go as far as setting education targets based on international assessments. For instance, Mexico established as an educational target in its 2007-12 Education Sector Programme, a combined score of 435 in the reading and mathematics PISA tests to be attained by 2012 (SEP and INEE, 2012). Broadfoot and Black (2004) note how assessment has become an international field: "Not only are new assessment policies and practices rapidly exported around the world, an increasing volume of assessment activity is explicitly international in being designed to compare national indicators and performance." They highlight three key themes: first, the increasing willingness of researchers to acknowledge the impact of context in the operation of particular assessment practices; second, the global scale and impact of assessment policy and practice; and third, the development of international surveys of learning and achievement.

The expansion of international assessment has also significantly contributed for some countries to introduce national standardised assessments. This was the case, for example, in Denmark, Italy, Mexico and Portugal, where there previously had been little emphasis on the measurement of student outcomes. In these countries, measured standardised student outcomes were only available through international assessments and it was deemed necessary to develop measures aligned with national student learning objectives. In some countries, national assessments were developed on the basis of methodology used by international assessments, particularly the assessment framework proposed by PISA. This has also translated into the exchange of expertise on the measurement of learning outcomes across countries.

Greater technological sophistication

The expansion of assessment, particularly the spreading out of standardised student assessment, as well as the management of the data it generates has greatly benefited from greater capacity of information and communication technologies. Improvements include more individualised assessment approaches, better assessment of cognitive skills such as

problem solving, capacity for rapidly marking large-scale assessments, reliability in marking and reduced cost to administer student assessment. For instance, in Denmark, computer-based national tests officially implemented in 2010 are adaptive in that the items are tailored to students' latent ability levels. Test items are selected sequentially according to a student's performance on the previous test items. These efficient national tests provide rapid feedback of test results to teachers the next day, which can greatly facilitate teachers' use of the test results (Danish Ministry of Education and Rambøll, 2012). In Norway, as of 2008, all primary and secondary schools were using an electronic test administration system and an electronic test execution system. It became possible to give ICT-based examinations for anyone interested, including for students sitting the national standardised tests (Norwegian Directorate for Education and Training, 2011). Other examples include the development of rapid-assessment – a computer-facilitated approach to frequent, brief formative student assessment, more sophisticated value-added models to determine a school's or a teacher's contribution to student learning, and data information systems providing new opportunities for information sharing across school agents.

Contextual developments shaping evaluation and assessment

Changing modes of school governance

Greater decentralisation and school autonomy

There is an increased prominence of evaluation and assessment as school systems decentralise with further autonomy given to intermediate levels of government (e.g. regions, municipalities) and to individual schools. There has been a general international trend towards devolution of responsibilities for budget management, staffing, educational provision, teaching content and processes, and the organisation of learning to the local level including schools. This increased autonomy has been balanced by the strengthening of accountability requirements for local education authorities and schools. This goes alongside the increasing role of central authorities in areas such as strategic steering, standard setting, support and capacity development. According to Eurydice (2007), in Europe, the expansion of school autonomy occurred mostly in the 1990s (e.g. Nordic countries, Austria, Czech Republic, Estonia, Hungary, Italy, Poland, Slovak Republic, Scotland) even if some countries pioneered it prior to the 1990s (e.g. Belgium, England, France, the Netherlands, Northern Ireland, Spain, Wales) and expansion continued in the 2000s in a more limited extent in some countries (e.g. Germany, Luxembourg, Portugal). However, it should be noted that the concepts of school autonomy differ considerably across countries. According to OECD (2012), in lower secondary education, the percentage of decisions taken at either the local or the school level exceeds 75% in the following systems: Chile, the Czech Republic, Denmark, England, Estonia, Finland, Hungary,[1] Iceland, the Netherlands, New Zealand, Norway, Scotland, Sweden and the United States.

In a context of school autonomy, greater policy attention is given to areas such as school leadership, capacity for schools to self-manage (including self-evaluation and the monitoring of the quality of teaching and learning) and ability to implement improvement processes. In addition, the greater responsibilities assumed by schools imply greater accountability requirements such as external school evaluation and public reporting of student performance.

Stronger role for market-type mechanisms

Another major trend in some countries is the growing use of market-type mechanisms in education to generate efficiencies and improve the quality of education systems. A market mechanism in education is an instrument that facilitates the co-ordination between the demand for and the supply of education services. The rationale for the introduction of market mechanisms is the expectation they will generate better outcomes in education systems than traditional regulatory instruments. Examples of market mechanisms are parental choice of schools (i.e. the ability for parents to channel public subsidies to the school of their choice, possibly a private school) and performance-based rewards or sanctions for schools and teachers. Both these mechanisms encourage competition among schools. In this context, school autonomy is seen as providing the latitude for the school to devise particular strategies to compete with other schools and demonstrate high performance publicly.

Several countries have attempted to raise educational quality by enhancing parental choice and allowing schools to compete for students. In Chile, the market-oriented education reforms of the 1980s entailed the decentralisation of public school management responsibilities to municipalities and the introduction of a nationwide voucher programme (Cox, 2005). The latter is characterised by a flat per student public subsidy for schools which are part of the voucher system (municipal schools and the majority of private schools), complemented with schemes to provide extra funding on the basis of educational disadvantage, and parents' free choice of schools. Other countries where parental choice of schools is extensive include Belgium, Ireland, the Netherlands and Sweden. In a school system significantly relying on parental choice and competition, evaluation serves two main purposes: to assure that schools are meeting the centrally defined requirements that justify their receipt of public funds; and to assure that parents have reliable information to assist with their decisions. Information about the quality of education services provided by schools (e.g. publication of student results at the school level, publication of school evaluation reports) is essential to achieve these purposes.

Some countries have also strengthened performance-based incentives for schools. For instance, in Mexico, the Incentives Programme for Teacher Quality, introduced in 2008, provides collective and individual stimuli to teachers and school leaders working in schools obtaining the highest results or most significant progressions in national student assessments (Santiago et al., 2012b). Performance-based incentives as a market mechanism require elaborate evaluation and assessment procedures to determine performance levels.

The emergence of New Public Management

The expansion of evaluation and assessment within education systems also reflect governments' efforts to "modernise" the public sector and incorporate business practices into public service management. This trend, often referred to as New Public Management (NPM) or Results-Based Management, aims to reform public sector operations by improving cost-effectiveness, measuring output and making public bodies with greater autonomy accountable to citizens and system managers (Mons, 2009). It involves greater emphasis on quality assurance and quality management in the public sector. NPM puts emphasis on leadership principles, incentives and competition between public sector agencies and private entities to enhance the outcomes and cost-efficiency of public services (Parker and Gould, 1999).

Since the late 1980s, Sweden undertook far-reaching public sector reforms to ensure a more efficient government administration. In the education sector, this led to the introduction of a system of *management by objectives*, which underlies all educational activities, including evaluation and assessment. The purpose of management by objectives is to increase efficiency in central administration by setting goals and assessing outcomes rather than focussing on input and processes (Nusche et al., 2011b). As described in Burgess and Ratto (2003), the use of explicit incentives to improve the efficiency of the public sector was an important component of the United Kingdom's public-service modernisation agenda of the late 1990s. The White Paper "Modernising Government" of 1999 emphasised the role of financial and other incentives in promoting better performance, leading to the development of performance indicators and systems of measuring and monitoring performance.

The rising importance of education in a global world

Economic activity has become globally interconnected on an unprecedented scale. The global character of markets has become stronger through international agreements and technological advances that bring people, goods and services together ever more quickly and less expensively. This growing integration of economies has an impact on strategies for national competitiveness, innovation, employment and skills (OECD, 2013). The emergence of the "knowledge society" and the strong skill bias in technological change have increased the value of education as a determinant of social and economic outcomes; this raises the payoff to good performance and amplifies the penalty for poor performance (OECD, 2008). The quality of education is necessary to achieve economic competitiveness in a context of global economic competition.

As a result, many of the proponents of national and international assessment place assessment in the context of a global economy, particularly in the context of being able to define a country's position in educational achievement relative to that of economic competitors, on the assumption that performance on measures of scholastic achievement has implications for economic performance (Kellaghan and Greaney, 2001). International student assessments have, to some extent, become a measure of a country's ability to compete in the global market and to drive economic growth. This has contributed to the expansion of evaluation and assessment activities in countries.

The growing imperative of an efficient use of public resources

The efficient use of resources is a growing concern. Education is costly and getting more so. OECD countries spent on average 6.2% of GDP on education institutions in 2009; between 1995 and 2009 the education share of public expenditure increased from 11.7% to 13.0% and real expenditure per student in pre-tertiary education increased by 55% between 1995 and 2009 (OECD, 2012).

The current financial crisis has intensified the need for efficiency in the use of public funds for education. Countries such as Greece, Ireland, Italy, Portugal and Spain face severe austerity measures which include cuts in the public education budget typically involving salary cuts for personnel working in public education, the freezing of career progression in the public service, and administration for education downsized. For example, in Ireland, the *value for money* imperative has been a fundamental part of public service modernisation and this has given an additional importance to evaluation and assessment in the educational context (Irish Department of Education and Skills, 2012). With challenging financial circumstances, the emphasis is on achieving greater efficiency

from the expenditure base. Effective monitoring, appraisal and evaluation is regarded as critical for delivering on this objective with a focus on the delivery of outputs and the achievement of goals/objectives. In Ireland *value for money* is a clear part of the rationale underpinning the National Literacy and Numeracy Strategy and the actions that have been identified to implement this plan (Irish Department of Education and Skills, 2012). Evaluation and quality assurance in education have become a necessity for policy makers to demonstrate that public funds are spent effectively and that the public purposes for financing education are actually fulfilled.

A shift from quantity to "quality for all"

In the great majority of OECD countries, attendance of lower secondary education is nearly universal – enrolment rates of 5-14 year-olds reached 95.9% on average across the OECD in 2010 (OECD, 2012) –, and graduation rates from upper secondary education have considerably increased – they reached 84% on average in the OECD area in 2010 (OECD, 2012). This is shifting the attention of policy makers from quantitative expansion to achieving "high-quality education for all". With the move towards knowledge-driven economies and societies, education has never been more important for the future economic performance and relative economic standing of countries, but also to allow individuals to perform and fully participate in the economy and society (OECD, 2007a). In this context, broad participation in education is only one side of the coin. The quality of education delivered is important to ensure that school graduates are effectively equipped to participate in the new economy and society at large, capable to learn at a higher level, and prepared to subsequently engage in lifelong learning activities to update their knowledge and skills. As a result, the issue of quality provision has received more and more interest from the various stakeholders over the past few decades. The greater stress on quality has given more prominence to evaluation and assessment activities.

Well-designed evaluation and assessment activities are expected to ensure that: each student is provided with quality and relevant education; the overall education system is contributing to the social and economic development of the country; and each school agent is performing at their best to deliver efficient education services. A corollary of this is that educational goals place increasing emphasis on equity objectives, which enlarges the scope for evaluation and assessment activities.

Rising expectations of the professionalism of teachers

The quality of learning and the successful implementation of education reforms depend crucially on teachers who are facing rising demands (OECD, 2005). The more complex and uncertain the world in which we live, the more that alternative sources of knowledge and influence are available to students, the more open schools become to diverse clienteles, and the more varied the organisational and pedagogical strategies that teachers should deploy, the greater become the levels of professional skill needed to meet them. There are growing expectations that teachers can operate in new organisational structures, in collaboration with colleagues and through networks, and be able to foster individual student learning. These call for demanding concepts of professionalism: the teacher as facilitator and knowledgeable, expert individual and networked team participant, oriented to individual needs and to the broader environment, engaged in teaching and in research and development (OECD, 2001).

This has implications for the evaluation and assessment framework. First, the standards by which teachers are appraised need to reflect the increasingly demanding

definitions of teacher professionalism. Second, teacher appraisal assumes a key role in identifying professional development needs in the process of acquiring the wider range of skills and competencies needed to meet professional expectations. Third, teacher professionalism is expected to be central in the effective implementation of evaluation and assessment policies through their understanding of evaluation and assessment procedures as well as their commitment to them.

More educated parents

The rising general education attainments of the population have a range of impacts on the world of the school, particularly of reducing the distance between schools and teachers, on the one hand, and the general public and parents, on the other. Many are now very familiar with the world of education, and are themselves qualified to levels at or greater than teachers (OECD, 2001). A result is that parents and others are more articulate and more demanding of the work of schools and teachers. This contributes to pressures for greater accountability in education, requires schools to become more transparent, and leads education systems to be more demand-driven (OECD, 2006).

Klenowski and Wyatt-Smith (2012) describe, in relation to the publication of the results of the National Assessment Program – Literacy and Numeracy (NAPLAN) in Australia, the federal government reporting of the high levels of parental support for the initiative, indicating that it believed that it serves the best interests of transparency and accountability. Parents are also gaining greater voice as countries extend opportunities for parental and student feedback through questionnaires and surveys conducted at the school, national and international levels, which assess their levels of satisfaction across a range of educational areas.

Greater sophistication of systems for the certification of learning and the recognition of competencies

In recent years, a number of countries have developed sophisticated national qualification frameworks in view of certifying learning and recognising competencies in education settings providing increasing flexibility for the individual to define his or her own learning pathway. Modularisation and credit systems, as well as qualification frameworks with several qualification levels, grant individuals with considerable flexibility in their learning, particularly at the upper secondary level within the formal school system (Dufaux, 2012). In New Zealand, for example, an elaborated qualification framework has been put in place to enable students to individualise their learning and have it formally recognised. The main qualification in secondary education is the National Certificate of Educational Achievement (NCEA), in which students are assessed against a range of National Standards. These consist of over 26 000 unit standards (vocationally based and mostly used in workplace training and the tertiary sector) and about 850 achievement standards (academically based and focused on the secondary school curriculum). Schools can design and offer their own courses mixing unit standards and achievement standards. This allows students to choose their personal learning content (Nusche et al., 2012).

The greater sophistication of certification systems, including the individualisation of learning, brings new challenges to student assessment. It requires flexible assessment instruments, accuracy in the assessment of very specific standards, and high capacity to assess and administer the qualifications framework. In these contexts, countries use a

greater variety of assessment tools such as short tests, projects, field studies, practical and aural tests (see also Chapter 4).

The growing importance of evidence-based policy

Another driver of the expansion of evaluation and assessment in school systems is the wider movement both within and beyond education to advance the use of evidence-based decision making. Data on student achievement are increasingly being used to support effective policy and practice, and to move education systems towards more evidence-informed approaches to large-scale improvement (Campbell and Levin, 2009). The public, professionals and policy makers want to know that their decisions, investments and actions are based on evidence. In education, this requires balancing schools' needs for data with external requirements and reporting (Campbell and Levin, 2009).

According to an OECD study on evidence in education (OECD, 2007b), the recent resurgence of interest in evidence-informed policy research can be explained by a range of factors such as: a greater concern with student achievement outcomes; a related explosion of available evidence due to a greater emphasis on testing and assessment; more explicit and vocal dissatisfaction with education systems, nationally and locally; increased access to information via the Internet and other technologies; and resulting changes in policy decision making. This highlights the interconnection between evidence-based policy and evaluation and assessment in education systems. Evaluations and assessments are key elements in the decision-making process. They provide the information on which accountability judgements are made and the means for steering improvement in educational practice.

Technological advancements

Information technology has developed very rapidly over the past 40 years, with computers becoming smaller, faster, cheaper, and more powerful. The ease and speed at which very large quantities of information can be rapidly accessed in a variety of settings have considerably improved (OECD, 2013). The digital revolution has drastically improved capacity to store, transmit, access and use information. The cost of transmitting information has significantly fallen, leading to the quasi abolition of physical distance. This has led to new developments in education technology – from Internet access to new teaching techniques enabled by classroom computers, which are driving changes in the education environment (ECS, 1999).

Countries are making significant investment in educational ICT infrastructure and equipment, and technology is increasingly being used to change what happens in the classroom and the school. ICT offers many opportunities to store and share data, to manage large amounts of information, to foster dialogue among education professionals, to strengthen feedback mechanisms and to improve the sophistication of evaluation and assessment procedures. The growing volume of data at all levels – student, teacher, school, local, national, and international – on education inputs and outcomes makes the monitoring of performance much easier, almost in real-time. Easier forms of communication (e.g. e-mail) improve the involvement of parents in school and their interest in following their children's progress. At the same time, teachers are able to use technology for professional development, online research and classroom and administrative data gathering (ECS, 1999). This is in addition to improvements in system-level initiatives such as the design, implementation and scoring of student standardised assessment; the development of Internet platforms to share education data among

stakeholders; and data information systems to facilitate knowledge management within evaluation and assessment frameworks.

The emergent commercial interests in education

The private sector is more and more a large player in the provision of ancillary services in education. In most OECD countries, it typically provides services such as the design and implementation of student standardised assessment, student private tutoring, online educational materials for students and teachers, textbooks and resources for school or classroom management. This is extensively the case in the United States. According to Burch (2009), the No Child Left Behind (NCLB) Act has helped private firms enter local education markets. She argues that the firms draw on political networks, new technologies and capital investments to become major suppliers to school systems for a vast array of educational services, including test score data storage, remedial instruction for the poor, online curriculum and online school management. She further claims that, triggered by high-stakes accountability policies, companies have introduced products and services that elevate the importance of standardised student assessments, private tutoring and technologies for school management (Burch, 2009).

In some countries, much of the activity of firms has concentrated on standardised student assessment, which is a growing and profitable industry. For instance, in the United States, the NCLB Act requires approximately 45 million standardised tests annually with considerable associated costs for developing, administering, publishing, scoring and reporting NCLB standardised tests (Toch, 2006). Burch (2009) emphasises that the market for test development and preparation has exploded in recent years. According to her, in 2006, the top vendors reported annual sales in the range of USD 100-600 million, with a pattern of increasing sales since the adoption of NCLB. Moreover, the testing market in the United States is dominated by only a handful of companies, which represent 90% of testing revenue (Toch, 2006). As standardised student assessment becomes a more profitable industry, companies have strong incentives to lobby for the expansion of student standardised assessment as an educational policy therefore influencing the activities within the evaluation and assessment framework.

The media as a driver of accountability in education

An important contextual influence for the development of evaluation and assessment is the role of the media in education. As data on student performance becomes readily available, as there is growing pressure for an effective use of public funds, and as the general public demands transparency in the delivery of education services, the media increasingly engages in the education public debate and makes information about student performance available, particularly school league tables. Governments are under pressure to release such information, also to prevent misinterpretation of data often presented by newspapers in simplistic ways. In addition, the attention the media devotes to education issues is also an opportunity for governments to publicise their accomplishments and feature the impact of their education policies. Given the greater mediatic impact of accountability policies, particularly those involving the measurement of student outcomes which have the potential of being reflected in school league tables, there are risks that accountability in education is, to some extent, driven by pressures from the media.

Notes

1. After two decades of decentralisation, Hungary has experienced a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities (including funding allocation) from 2013 onwards. Except when explicitly indicated, information about Hungary in this report refers to the period prior to this reform.

References

Bonnet, G. (2004), "Evaluation of education in the European Union: Policy and methodology", *Assessment in Education*, Vol. 11, No. 2.

Broadfoot, P. and P. Black (2004), "Redefining assessment? The first ten years of *Assessment in Education*", *Assessment in Education*, Vol. 11, No. 1.

Burch, P. (2009), *Hidden Markets: The New Education Privatization*, Taylor and Francis.

Burgess, S. and M. Ratto (2003), "The role of incentives in the public sector: Issues and evidence", *Oxford Review of Economic Policy*, Vol. 19, No. 2.

Campbell, C. and B. Levin (2009), "Using data to support educational improvement", *Educational Assessment, Evaluation and Accountability*, No. 21, pp. 47-65.

Cox, C. (ed.) (2005), *Educational Policies for the Turn of the Century: The Chilean School Reform*, Editorial Universitaria, Santiago, Chile.

Danish Ministry of Education and Rambøll (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Update to the Country Background Report for Denmark*, National Agency for Quality and Supervision, *www.oecd.org/edu/evaluationpolicy*.

Dos Santos, S. and R. Rakocevic (2012), *Les cadres d'évaluation en vue d'améliorer les résultats scolaires : Étude de l'OCDE – Rapport national de base pour la France.*

Dufaux, S. (2012), "Assessment for qualification and certification in upper secondary education: A review of country practices and research evidence", *OECD Education Working Papers*, No. 83, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

ECS (Education Commission of the States) (1999), *Future Trends Affecting Education*, Denver, Colorado, *www.ecs.org/clearinghouse/13/27/1327.htm*.

Elmore, R.F. (2004), "The problem of stakes in performance-based accountability systems", in S.H. Fuhrman and R.F. Elmore (eds.), *Redesigning Accountability Systems for Education*, Teachers College Press, New York.

Eurydice (2009), *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*, Eurydice, Brussels.

Eurydice (2007), *School Autonomy in Europe: Policies and Measures*, Eurydice, Brussels.

Figlio, D. (2006), "Testing, crime and punishment", *Journal of Public Economics*, Vol. 90, pp. 837-851.

Figlio, D. and L. Kenny (2009), "Public sector performance measurement and stakeholder support", *Journal of Public Economics*, Vol. 93, Issues 9–10, pp. 1069–1077.

Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, *www.oecd.org/edu/evaluationpolicy*.

Hamilton, L. (2003), "Assessment as a policy tool", *Review of Research in Education*, Vol. 27 (1), January, pp. 25-68.

Hamilton, L. and D. Koretz (2002), "Tests and their use in test-based accountability systems", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-based Accountability in Education*, RAND, Santa Monica, California, pp. 13-49.

Hargreaves, A. and D. Shirley (2009), *The Fourth Way: The Inspiring Future for Educational Change*, Corwin, Thousand Oaks, California.

Hout, M. and S. Elliott (2011), *Incentives and Test-Based Accountability in Education*, Committee on Incentives and Test-Based Accountability in Public Education, National Research Council, The National Academies Press, Washington, D.C.

Irish Department of Education and Skills (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Ireland, www.oecd.org/edu/evaluationpolicy*.

Kellaghan, T. and V. Greaney (2001), "The globalisation of assessment in the 20th century", *Assessment in Education*, Vol. 8, No. 1.

Klenowski, V. and C. Wyatt-Smith (2012), "The impact of high stakes testing: The Australian story", *Assessment in Education: Principles, Policy and Practice*, Vol. 19, No. 1, February, pp. 65-79.

Looney, J. (2009), "Assessment and innovation in education", *OECD Education Working Papers*, No. 24, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Mons, N. (2009), "Theoretical and real effects of standardised assessment", background paper to the study: National Testing of Pupils in Europe, Eurydice Network, *http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111EN.pdf*, accessed 12 July 2011.

New Zealand Ministry of Education (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for New Zealand, www.oecd.org/edu/evaluationpolicy*.

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

OECD (2013), *Trends Shaping Education 2013*, Centre for Educational Research and Innovation, OECD Publishing, Paris.

OECD (2012), *Education at a Glance: OECD Indicators 2012*, OECD Publishing, Paris.

OECD (2011), *Education at a Glance: OECD Indicators 2011*, OECD Publishing, Paris.

OECD (2008), "How do OECD countries take stock of progress and performance in education systems? Evidence and issues", paper prepared for the Education Policy Committee Ad Hoc Workshop "Taking Stock of Education Performance: From Student Testing to System Evaluation", EDU/EDPC(2008)33, Paris.

OECD (2007a), *Human Capital: How What You Know Shapes Your Life*, OECD Publishing, Paris.

OECD (2007b), *Evidence in Education: Linking Research and Policy*, OECD Publishing, Paris.

OECD (2006), *Schooling for Tomorrow: Demand-Sensitive Schooling? Evidence and Issues*, Centre for Educational Research and Innovation, OECD Publishing, Paris.

OECD (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OECD Publishing, Paris.

OECD (2001), *Schooling for Tomorrow: What Schools for the Future?* Centre for Educational Research and Innovation, OECD Publishing, Paris.

Parker, L. and G. Gould (1999), "Changing public sector accountability: Critiquing new directions", *Accounting Forum*, Vol. 23, No. 2.

Roeber, E. (1988), "A history of large-scale testing activities at the state level", paper presented at the Indiana Governor's Symposium on ISTEP, Madison.

Santiago, P., F. Benavides, C. Danielson, L. Goe and D. Nusche (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012b), *OECD Reviews of Evaluation and Assessment in Education: Mexico*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Schildkamp, K., M. Ehren and M.K. Lai (2012), "Editorial article for the special issue on data-based decision making around the world: From policy to practice to results", *School Effectiveness and School Improvement*, Vol. 23, No. 2, June, pp. 123-131.

SEP and INEE (Secretariat of Public Education and National Institute for Educational Assessment and Evaluation) (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Mexico*, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011a), *OECD Reviews of Evaluation and Assessment in Education: Denmark,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., M. Hulshof, D. Nusche and L. Stoll (2011b), *OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Toch, T. (2006), *Margins of Error: The Education Testing Industry in the No Child Left Behind Era,* Education Sector Reports, *www.educationsector.org/publications/margins-error-testing-industry-no-child-left-behind-era*, accessed 12 July 2011.

Wang, L., G. Beckett and L. Brown (2006), "Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence", *Applied Measurement in Education*, Vol. 19, No. 4, pp. 305-328.

Chapter 3

The evaluation and assessment framework:

Embracing a holistic approach

This chapter looks at the overall framework for evaluation and assessment, i.e. its various components such as student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation, the coherence of the whole as well as the articulation between the different components. This chapter supports the view that evaluation and assessment in school systems need to be conceived holistically, as a whole framework, building on the interdependence of its parts in order to generate complementarities, avoid duplication, and prevent inconsistency of objectives. The chapter illustrates the synergies that can be generated between the different evaluation and assessment components. Areas analysed include the governance and structure of the framework, capacity for evaluation and assessment, alignment with educational goals, articulations within the framework, knowledge management, the use of evidence to inform policy development and strategies in the implementation of evaluation and assessment policies.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter looks at the overall framework for evaluation and assessment, i.e. its various components such as student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation, the coherence of the whole as well as the articulation between the different components. Most OECD countries have not conceived evaluation and assessment as an integrated framework but instead developed a number of independent components operating at different levels. This chapter supports the view that evaluation and assessment in school systems need to be conceived holistically, as a whole framework, building on the interdependence of its parts in order to generate complementarities, avoid duplication, and prevent inconsistency of objectives. The chapter illustrates the synergies that can be generated between the different evaluation and assessment components in response to the overarching policy question of the OECD Review: "How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools?"

Evaluation and assessment may operate at six key levels: national education system, education sub-systems (e.g. states), local education authority, school, teacher and student. At each of these levels, evaluation and assessment mechanisms provide a basis for assessing how effectively education is being provided for students. The ultimate objective is to improve the quality of education in countries and, as a consequence, raise student outcomes. Evaluation and assessment typically inform policy development, curriculum, planning, reporting, improvement strategies, budgetary choices, resource allocation decisions, and performance management.

This chapter reviews the main features of evaluation and assessment frameworks and countries' overall approaches to evaluation and assessment. The chapter begins by reviewing concepts and the several dimensions of an evaluation and assessment framework. It also examines country practices across the different dimensions of the framework. Areas analysed include the governance and structure of the framework, capacity for evaluation and assessment, alignment with educational goals, articulations within the framework, knowledge management and the use of evidence to inform policy development. The chapter further examines factors and strategies in the implementation of evaluation and assessment polices. It concludes with a set of policy options for countries to consider.

Following this overview, the succeeding chapters will analyse the issues relevant to each individual component of the evaluation and assessment framework in more depth. An important consideration is to establish the distinction between the evaluation and assessment framework (this chapter) and the evaluation of the education system (Chapter 8). While the evaluation of the education system focuses specifically on evaluation procedures to assess the extent to which objectives of education systems are achieved, the evaluation and assessment framework deals with the systemic governance of the whole range of evaluation and assessment activities in a school system, including strategies for the various components to complement each other and articulate coherently to achieve given purposes.

Definitions and scope

Definitions

The *evaluation and assessment framework* consists of the co-ordinated arrangements for evaluation and assessment which ultimately seek to improve student outcomes within a school system. The framework typically contains various components as student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation, and includes the articulation between the components and their coherent alignment to student learning objectives.

This framework differentiates between the terms *assessment*, *appraisal* and *evaluation*:

- The term ***assessment*** is used to refer to judgements on individual student progress and achievement of learning goals. It covers classroom-based assessments as well as large-scale, external assessments and examinations.
- The term ***appraisal*** is used to refer to judgements on the performance of school-level professionals, e.g. teachers, school leaders.
- The term ***evaluation*** is used to refer to judgements on the effectiveness of schools, school systems, policies and programmes.

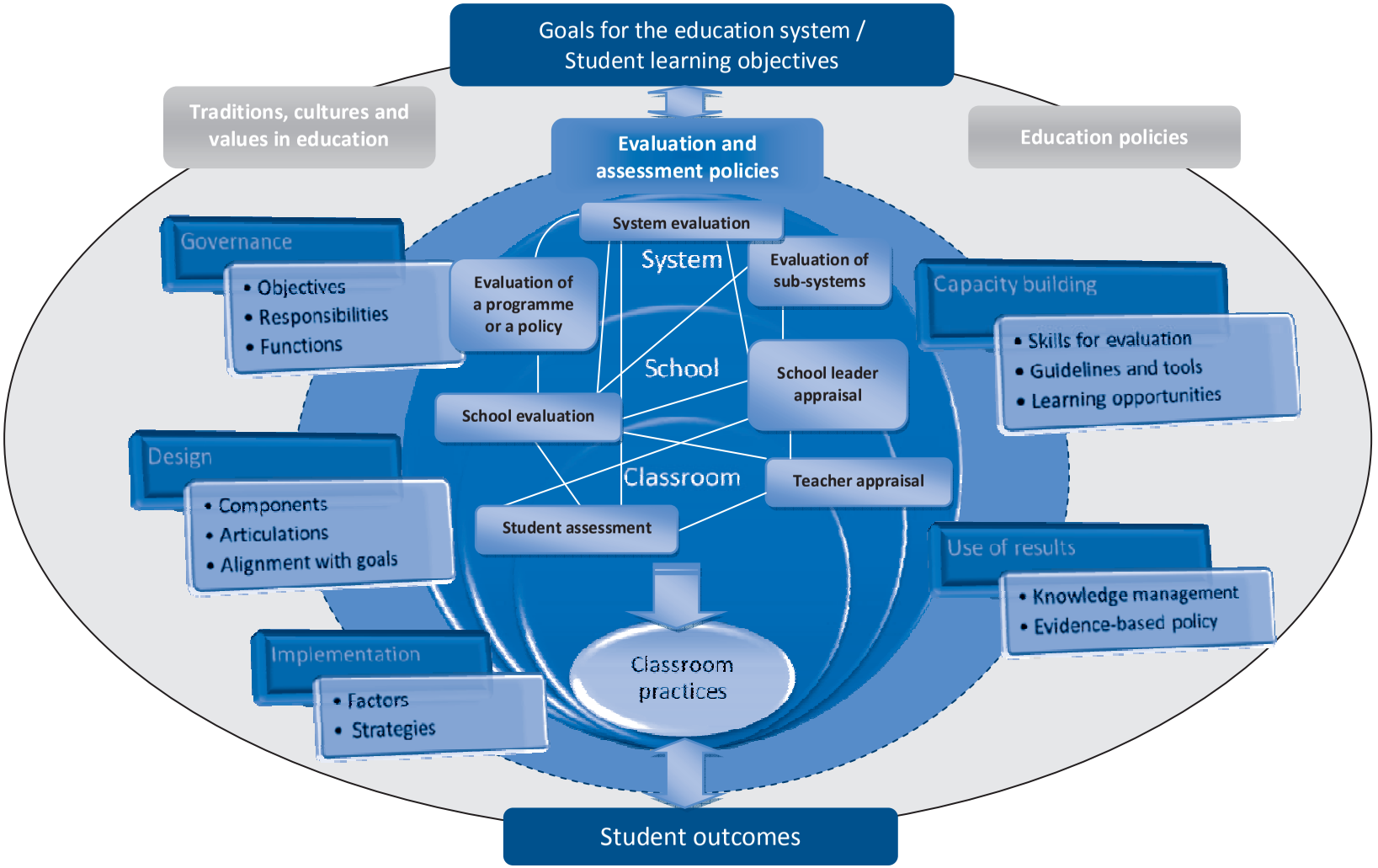
Scope

Figure 3.1 provides an overview summarising the main features of the overall evaluation and assessment framework. The areas covered are the following:

- **Governance** of the evaluation and assessment framework including objectives, distribution of responsibilities, functions within the framework, the concept of evaluation and assessment, and the integration of the non-public sector.
- **Design**. Configuration/architecture of the evaluation and assessment framework, including its main components, the articulation between them, the main principles on which evaluation and assessment procedures rely, alignment with educational goals, and the links to classroom practices.
- **Capacity building**. Competencies and skills for evaluation and assessment across the framework, including guidelines and tools for evaluation and assessment and learning opportunities.
- **Use of results** from the overall evaluation and assessment framework, including knowledge management and evidence-based policy.
- **Implementation** strategies and factors.

As conceived in Figure 3.1, the evaluation and assessment framework is part of the overall education system, is designed in alignment with the goals for the education system and seeks to improve student outcomes. It is also developed in the broader context set by the education policies prevailing in the concerned country (e.g. level of decentralisation of educational governance, degree of market mechanisms in education) as well as the country's traditions, cultures and values in education.

Figure 3.1 Main features of the overall evaluation and assessment framework



Educational context

The educational context shapes the evaluation and assessment framework

Evaluation and assessment frameworks are developed within the broader context of established education policies and existing traditions, cultures and values in education. Education policy dictates the need for and importance of evaluation and assessment while traditions and cultures in education shape the nature and significance of evaluation and assessment activities. Examples of important contextual aspects to the development of evaluation and assessment are:

- the culture of evaluation within the education system
- the tradition of quality-focussed policies in the education sector and within the public sector
- the extent to which evaluation and assessment is developed as a profession
- prevailing conceptions of evaluation and assessment
- extent to which teachers are trusted as professionals
- the level of decentralisation of educational governance and the extent of school autonomy
- prevailing approaches to school leadership
- the extent of the regulatory and standard-setting role of national education authorities
- the degree of market mechanisms in education (e.g. school choice, competition between schools, organisation of labour market for teachers)
- whether education staff are part of the civil service
- level of education of parents and their culture as "consumers"
- the financial circumstances of the public budget for education
- extent to which resources in education are efficiently used
- relative policy emphasis on quality versus quantity of education services
- capacity of information and communication technologies in the education system
- emphasis on evidence for education policy development
- the importance given to education by the media
- presence of commercial interests in education.

These have a strong impact on the design and implementation of evaluation and assessment policies. For instance, there is an increased prominence of external school evaluation in those countries where the devolution of responsibilities to the school level is extensive. The greater school autonomy is typically balanced with greater externality of evaluation. In this context, strong school leadership is needed to effectively exert autonomy and liaise with education authorities to meet accountability requirements.

In school systems significantly relying on parental school choice such as Chile, Iceland, the Netherlands and Poland, evaluation includes two purposes: to ensure that schools are meeting the centrally defined standards that justify their receipt of public funds; and to assure that parents have reliable information to assist with their choice of school. This means that in education systems emphasising parental school choice, the evaluation and assessment framework should give a key role to publicly available school-level information on student achievement. Also, a number of countries provide for considerable teacher accountability through market mechanisms. This is the case in Sweden where teacher pay is defined at the individual level by school leaders (within a municipal framework) and local authorities (municipalities) or schools have full autonomy in hiring their teachers. In addition, the "funding follows the student" as when a student moves school, the operating grant that applies to that student is reallocated to their new school (regardless of it being a municipal or an independent school). This leads to strong competitive pressures on schools and teachers which condition, for example, approaches to teacher appraisal (Nusche et al., 2011a). These effects are reinforced in those countries where there is a strong tradition of parents and students as consumers of education services.

In Belgium, the constitutional Freedom of education has a wide impact on evaluation and assessment frameworks in both the Flemish and the French Communities. Freedom of education has a twofold implication: *(i)* every natural or legal person is free to start a school (freedom of organisation); and *(ii)* every parent can choose freely in which school to enrol his or her child. The "freedom of organisation" considerably shapes the evaluation and assessment framework as schools are free to develop their own approaches to education quality within the boundaries set by the regulatory framework. As a result, schools are the key actors in evaluation and assessment, but need to follow the requirements set forth by the evaluation and assessment framework (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010, and Blondin and Giot, 2011).

New Zealand has developed its own distinctive model of evaluation and assessment that is characterised by a remarkable level of trust in schools and school professionals. New Zealand's approach relies on national standard setting and test development combined with strong school autonomy in implementing evaluation and assessment. The education system aims to make the best use of student achievement data to inform decision making at all levels while limiting possible negative impacts of high-stakes assessment. There is a general consensus against national testing and a strong opposition to the use of student data for comparison among schools, such as league tables, especially in primary education (Nusche et al., 2012).

Governance

Objectives and functions of evaluation and assessment

Improving student outcomes is the ultimate objective of evaluation and assessment

The ultimate objective of evaluation and assessment is to enhance student outcomes through the improvement of practices at the different levels of the school system, including teaching methods, school leadership processes, ways to organise learning, and directions of education policy.

Some education systems explicitly define the objectives of educational evaluation. In Finland, a specific decree on the evaluation of education sets out the following aims for evaluation: to provide and analyse data in support of national decision making on education and as a basis for educational development; to provide and analyse data as a basis for local educational development and decision making; and to support learning, the work of school personnel and institutional development (Finnish Ministry of Education and Culture, forthcoming). In Portugal, the objectives for the evaluation system are legally defined by the Basic Law on the Education System, which positions it as the central instrument for developing education policies. The objectives of the evaluation system are defined as: *(i)* promoting improvement in the quality of the education system, including in its organisation, efficiency and efficacy levels; *(ii)* supporting the formulation, development and implementation of education and training policies; and *(iii)* ensuring the availability of information about the system for its management (Santiago et al., 2012a).

Evaluation and assessment have a range of functions

Evaluation and assessment provide a basis for monitoring how effectively education is being delivered to students and for assessing the performance of systems, schools, school leaders, teachers and students, among others. They can serve as an instrument for the *accountability* of school agents when the results of an evaluation or assessment have stakes for school agents such as linkages to career advancement or salary progression, one-off rewards, sanctions, or simply information to parents in systems based on parental school choice. By measuring student outcomes and holding teachers and schools responsible for results, accountability systems intend to create incentives for improved performance and identify "underperforming" schools and school agents. Evaluation and assessment also identify strengths and weaknesses of systems, schools, school leaders, teachers and students which inform areas for *development*. In addition, evaluation and assessment can have a *diagnostic* function such as with school readiness assessments or sampled-based standardised assessments to measure the extent to which student learning objectives are being achieved across the education system.

The balance between the development and the accountability functions varies across countries

An important characteristic of evaluation and assessment frameworks is the balance between the accountability and the development functions. Some countries emphasise one function over the other. Countries with a strong focus on accountability typically emphasise: high-stakes standardised assessment of students; teacher appraisal that is linked to decisions regarding career advancement, salary, promotion and dismissal; external reviews or inspections of school quality; publication of school evaluation results and/or public comparisons of school performance. Countries with a strong focus on development and improvement typically emphasise: formative, low-stakes assessment of students; teacher appraisal that is linked to decisions regarding teacher professional development and learning opportunities; school self-evaluation and external support for organisational learning.

In Mexico it is apparent that the policy initiatives in evaluation and assessment of the last few years have emphasised accountability over development. For instance, the in-service teacher appraisal system currently in place is mostly focussed on salary progression and rewards and places less emphasis on its links to professional development, the assessment of students is oriented towards summative scores, and school evaluation is

essentially reduced to accountability through the publication of student standardised assessment results (Santiago et al., 2012b). By contrast, in New Zealand, the development function of evaluation and assessment is strongly emphasised. In primary schools, student assessment is mostly formative and provides detailed feedback rather than assigning numerical marks. The sample-based standardised assessments do not carry high stakes for students. Assessment in secondary schools is more summative but there are opportunities for schools to reassess and resubmit internal assessments to maximise learner success and students also receive their marked external assessments back. A range of tools and professional development offers are available for teachers to help them gather a variety of evidence of student learning to allow nuanced overall judgements on performance. Also, New Zealand's approach to school evaluation has evolved to focus attention on building the capacity of schools for effective self-evaluation and strategic planning for improvement of teaching and learning (Nusche et al., 2012).

Table 3.1 provides a taxonomy of country approaches to the use of results for accountability and for development. Countries are grouped according to two dimensions: the extent of use for accountability and the extent of use for development. The taxonomy uses information supplied by countries on features of evaluation and assessment frameworks (synthesised in Chapters 4 to 8). The assessment depends on the degree of existence of systematic linkages to either development or accountability in each country. Some countries (e.g. Mexico, Slovak Republic) seem to give greater prominence to the accountability function while others (e.g. Denmark, Iceland, Norway) seem to place more emphasis on the development function of evaluation and assessment.

Table 3.1 Use of results for accountability and development across countries (2012)

Use of results for accountability	Use of results for development: High	Use of results for development: Moderate
High	Australia, Chile	Mexico, Slovak Republic, Sweden
Moderate	Belgium (Fl.), Canada, Israel, Korea, New Zealand	Czech Republic, France, Hungary, Ireland, Netherlands, Poland, Portugal
Low	Denmark, Iceland, Norway	Austria, Belgium (Fr.), Estonia, Finland, Italy, Luxembourg, Slovenia, Spain

Definitions: A country is considered as making a *high* use of the results of evaluation and assessment for development (accountability) if most of the components of its evaluation and assessment framework are systematically linked to actions for development (accountability); is considered as making a *moderate* use of such results if systematic linkages to actions for development (accountability) only exist for some of the components of the evaluation and assessment framework; and is considered as making a *low* use of such results if no systematic linkages to actions for development (accountability) exist for most or all components of the evaluation and assessment framework. This involved the computation of indices on the basis of the information supplied by countries on features of evaluation and assessment frameworks and information from *Education at a Glance 2011*. The index on "use of results for accountability" considered the accountability uses of teacher appraisal (impact on career advancement, existence of rewards or sanctions, impact on registration status), school evaluation (extent to which results are shared, financial implications, likelihood of school closure, rewards for school agents), school leader appraisal (impact on career advancement, existence of rewards or sanctions) and national student assessment (publication of results at the school level, links to school rewards/sanctions). The index on "use of results for development" considered the development uses of teacher appraisal (extent to which there are links to professional development), school leader appraisal (extent to which there are links to professional development) and national student assessment (whether or not results are shared with teachers).

Source: Taxonomy developed using information supplied by countries on features of evaluation and assessment frameworks (synthesised in Chapters 4 to 8) and information from *Education at a Glance* (OECD, 2011).

A key challenge therefore is to find the right balance between the accountability and the development functions of evaluation and assessment. While transparency of information, high-quality data, and the accountability of school agents are essential for a well-functioning evaluation and assessment system, it is important to ensure that the existing data and information are actually used for development and improvement. This requires reflection on designing mechanisms to ensure that the results of evaluation and assessment activities feed back into teaching and learning practices, school improvement, and education policy development.

There are often challenges in communicating the ultimate objective of evaluation and assessment

The idea that the ultimate objective of evaluation and assessment is to improve students' learning and teachers' teaching may not be adequately implemented, easily communicated or understood in countries' evaluation and assessment frameworks.

When the framework tends to stress the accountability function there is a risk that evaluation and assessment are perceived mostly as instruments to hold school agents accountable, to "control" and assess compliance with regulations. For instance, in the Czech Republic, this is visible at all levels with the focus often being whether formal requirements are met and with less attention given to the quality of practices or ways for these to improve. School inspections are much better established as an evaluation practice than school self-evaluation, which is not widespread and systematic across the system. Also, student assessment is perceived more as test and measurement rather than learning. This translates into a situation whereby the more accountability-oriented elements of the framework are receiving greater attention than processes for development, which risks leading to more limited local engagement in self-assessment activities, incipient practices of evidence-informed inquiry, and assessment and evaluation results not used to their potential (Santiago et al., 2012c).

Also, some countries emphasise the "measurement" dimension of evaluation and assessment. This is the case in Mexico through the dominance of national full-cohort standardised assessment (National Assessment of Academic Achievement in Schools, ENLACE). In addition to the primary purpose for which it was conceived, the formative assessment of students, ENLACE results constitute the dominant instrument in in-service teacher appraisal, the central factor in school accountability (through the publication of results at the school level), and the *de facto* key element in the evaluation of the national education system and the state education sub-systems. By contrast, qualitative evaluation, feedback for improvement and professional dialogue around evaluation results are all not sufficiently developed in the evaluation and assessment framework (Santiago et al., 2012b).

In other cases, the development function of evaluation and assessment is less well understood by school agents. In Norway, according to an evaluation of the National Quality Assessment System (NKVS), among stakeholders there is no clear understanding of the whole system for evaluation and assessment (Allerup et al., 2009). The evaluation showed that the key elements of NKVS were understood to be the national tests, user surveys, inspections and international tests. This reflects that the more accountability-oriented elements of the evaluation and assessment framework are more prevalent in the perceptions of school agents than the support and guidance tools developed by the Directorate for Education and Training for local use and analysis. This is in spite of the fact that the proposal for the creation of the Directorate had clearly stated that "quality

assessment should primarily be a tool to be used by teachers, schools and students in their quality development work" (Nusche et al., 2011b).

Hence, in some countries, there is a narrow understanding of the purposes and the potential of evaluation and assessment. The challenge is therefore to convey a more constructive view of evaluation and assessment, communicate that the ultimate objective of evaluation and assessment is to improve student learning and build an evaluation culture in the education system.

Responsibilities within the evaluation and assessment framework

Responsibilities for evaluation and assessment are shared among a wide range of agents

A wide range of agents take responsibilities within the evaluation and assessment framework, such as:

- education authorities at the national level (e.g. ministry of education, department of education)
- education authorities at the sub-national level (e.g. states, regions, municipalities)
- governing bodies in the non-public sector
- agencies for evaluation and assessment such as quality assurance agencies, inspectorates or school review agencies, and agencies overseeing strategies for educational development
- audit offices
- schools, school governing bodies and school leaders
- teachers' professional bodies
- teachers
- parents and students.

Most countries' approach to evaluation and assessment combines central direction (either at the national or sub-national level) over policy development and standard-setting with a measure of devolved responsibility for the implementation of evaluation and assessment at the local and school levels. The central direction often involves a number of agencies with key functions in the evaluation and assessment framework. Also, devolution of evaluation and assessment procedures to the local level typically comes along with national frameworks, guidance materials, and tools for the use of school agents.

Schools are also accountable to students and their parents, to members of the community, and to the community as a whole for multiple aspects of schooling, based on various information sources (Hooge et al., 2012). As explained in Hooge et al. (2012), with respect to multiple accountability processes, Hooge and Helderman (2008) distinguish four different categories of stakeholders: primary, internal, vertical and horizontal. In education, parents and students are the primary stakeholders. Teachers and other educational and non-educational staff are internal stakeholders with a clear interest in the success of the school. At slightly more distance, governments and organisations formally operating on behalf of government (such as inspectorates or municipalities)

operate as vertical stakeholders. Finally, all other organisations, groups, or persons in the school's environment with some level of interest in the school are horizontal stakeholders.

In Mexico, educational evaluation is a responsibility of federal education authorities. The Secretariat of Public Education (SEP) is responsible for the supervision of the entire education system and plays a role in all components of the evaluation and assessment framework, including developing binding student learning objectives (national curriculum), determining the features of the teaching profession, and monitoring the performance of schools and the education system. The SEP is supported by a federal-level agency, the National Institute for Educational Assessment and Evaluation (INEE). Activities of INEE include the design and development of student national assessments, educational indicators on the quality of education in Mexico, and evaluation instruments and guidelines (e.g. for school evaluation). State authorities operate schools in their sub-system. They organise their own systems of school supervision which tends to concentrate on the compliance with regulations and provide some support for schools to improve. In addition to the co-ordination of federal evaluation initiatives, states can also develop their own evaluations. Schools benefit from some limited autonomy in the organisation of the various components of evaluation and assessment at the student, teacher and school level. They take most responsibility for student assessment, including the definition of assessment criteria and instruments (mostly determined by individual teachers); they operate some elements of some teacher appraisal processes; and they take responsibility for their self-evaluation (SEP and INEE, 2011).

In Ireland, on behalf of the Minister, the Department of Education and Skills (DES) co-ordinates and develops policy and decisions relating to the monitoring and assessment of the education system taking into account advice from the NCCA (National Council for Curriculum and Assessment) which has a key advisory role in developing mechanisms for assessing standards of knowledge and skills. The Inspectorate of the DES is the key agency in evaluating the education system and in conducting school evaluation, which includes evaluation of teaching and learning in individual schools. The Inspectorate contributes regularly to system evaluation by undertaking evaluation of educational programmes designed to meet particular needs, by participating in Value-for-Money reviews or by assessing policy implementation and impact. The DES has devolved elements of student evaluation and assessment to national bodies or external agencies. The State examinations are co-ordinated and managed by the State Examinations Commission (SEC), a statutory independent agency. The national sampling assessment of students is carried out by the Educational Research Centre (ERC) which is an independent research agency (Irish Department of Education and Skills, 2012).

Denmark combines a central legal framework specifying evaluation requirements and student Common Objectives in compulsory education, with clear responsibility for school owners (municipalities and private school boards) to ensure quality control within this framework. Municipalities enjoy autonomy in designing their quality assurance practices, specifying the local objectives and determining local guidelines for their schools. School principals are responsible for school-level administrative and pedagogical policies and are accountable to the municipality (public schools) or the parent-elected boards and the Quality and Supervision Agency (private schools) (Danish Ministry of Education and Rambøll, 2011).

In Canada, the responsibilities for evaluation and assessment are shared between the education agents but the role of local players is prevalent. System evaluation is the responsibility of Provincial Departments of Education and local governing bodies (school

boards, groups of elected members of a community – school district – to whom the provinces have delegated authority over some aspects of education) and their supervisory staff (i.e. superintendents). School evaluations and associated activities are the responsibility of school staff and leadership, with measures included for review by supervisory staff of local governing bodies and reported to the appropriate board and the local public. Teacher appraisal is typically the responsibility of school principals with results reported to the supervisory personnel of local governing bodies, while student assessment is administered by a variety of players – teachers, local governing boards, and provincial authorities – depending on the intended uses of achievement results (Fournier and Mildon, forthcoming).

In the Czech Republic, the Ministry of Education, Youth and Sports (MEYS) oversees the entire education system and co-ordinates the evaluation and assessment framework, with a role in each of its components. In addition, the role of the Schools Inspectorate (CSI) in assuring quality in schools is perceived as central by all agents, some of whom considerably draw on its work (such as regions' and municipalities' education authorities). Regions and municipalities supervise their respective schools but in clear respect of school autonomy and also drawing on the framework provided by the national level, including the work of the CSI. It is also clearly understood that some areas within the evaluation and assessment framework are better addressed at the local level such as teacher appraisal and student assessment. In addition, school boards give an opportunity for parents to contribute to the evaluation and assessment framework (IIE, 2011).

Dedicated intermediate agencies gain a prominent role in evaluation and assessment frameworks

As described in Chapter 2, the greater emphasis on evaluation and assessment has led to the creation of specialised intermediate agencies which assume a central role in the governance of the evaluation and assessment framework. Table 3.2 provides a country-level list of agencies which play a role in the respective evaluation and assessment framework. Many of the listed agencies have been created in recent years in recognition of the increasing complexity of evaluation and assessment frameworks. These agencies are typically involved in the design and operation of evaluation activities (e.g. national standardised student assessments, external school evaluation), technical leadership (e.g. design of guidelines and instruments for evaluation), capacity building for evaluation and assessment and the monitoring of the education system.

The creation of some of these agencies was triggered by the establishment of national education standards and the need for their monitoring. For example, in Austria, the establishment of the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) was associated with the development, implementation and monitoring of education standards. Since its creation, BIFIE assumes responsibility for the preparation of the annual "National Education Report" (Specht and Sobanski, 2012). Similarly, in the Slovak Republic, the National Institute of Certified Measurement (NÚCEM) takes responsibility for measuring and evaluating the quality of education. In Italy, the National Institute for the Evaluation of the Education System (INVALSI) has been given the functional responsibility of system evaluation under the supervision of the National Ministry for Education, University and Research.

Table 3.2 Specialised intermediate agencies with a role in the evaluation and assessment framework, by country (2012)

Country	Specialised intermediate agencies with a role in the evaluation and assessment framework
Australia (at the federal level)	Australian Curriculum, Assessment and Reporting Authority (ACARA) (created in 2009) Australian Institute of Teaching and School Leadership (AITSL) (created in 2010)
Austria (at the federal level)	Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) (created in 2008)
Belgium (Flemish Community)	Agency for Quality Assurance in Education and Training (AKOV) (created in 2009) The Flemish Inspectorate of Education (created in 1991)
Belgium (French Community)	General Inspection Services (*Service général de l'inspection*) (associated with Ministry of the French Community) Commission for the monitoring of the education system (*Commission de pilotage*) (created in 2002)
Canada (at the federal level)	Council of Ministers of Education, Canada (Pan-Canadian Assessment Program)
Chile	The Quality of Education Agency (created in 2011, started operating in 2012)
Czech Republic	Czech Schools Inspectorate (created in 1991, reformed in 2005) Centre for the Evaluation of Educational Achievement (CERMAT) (under the Ministry of Education, Youth and Sports) (created in 1999)
Denmark	Quality and Supervision Agency (created in 2011) Council for Evaluation and Quality Development of Primary and Lower Secondary Education (created in 2006) Danish Evaluation Institute (created in 1999)
Finland	National Board of Education (created in 1991) Education Evaluation Council
France	General Inspectorate of National Education (part of the Ministry of Education) General Inspectorate of the Administration of National Education and Research (part of Ministry of Education)
Hungary	The Educational Authority (created in 2006)
Iceland	Educational Testing Institute
Ireland	Inspectorate of the Department of Education and Skills National Council for Curriculum and Assessment (NCCA) (created in 2001) State Examinations Commission (separate entity under Department of Education and Skills) (created in 2003) Educational Research Centre (created in 1966)
Italy	National Institute for the Evaluation of the Education System (INVALSI) (created in 2004) National Institute for Educational Research, Experimentation and Development (INDIRE)
Korea	Korea Institute for Curriculum and Evaluation (KICE) (created in 1998) Korean Educational Development Institute (KEDI) (created in 1972)
Luxembourg	Agency for the Development of Quality in Schools (integrated in the Ministry of National Education and Vocational Training) (created in 2009)
Mexico	National Institute for Educational Assessment and Evaluation (INEE) (created in 2002)
Netherlands	Dutch Inspectorate of Education Board of Examinations (CVE)
New Zealand	Education Review Office New Zealand Qualifications Authority (created in 1989)
Norway	Directorate for Education and Training (executive agency of the Ministry of Education and Research) (created in 2004)
Poland	Central Examination Commission
Portugal	General Inspectorate of Education and Science (integrated in the structure of the Ministry of Education and Science) (created in 1979)
Slovak Republic	National Institute for Certified Educational Measurement (NÚCEM) (created in 2008) State Schools Inspectorate
Slovenia	National Examinations Centre (created in 1993) Inspectorate of the Republic of Slovenia for Education and Sport (affiliated to the Ministry of Education and Sport) National Education Institute (created in 1956) Council for Quality and Evaluation (Advisory body of Ministry of Education and Sport, created in 1999)
Sweden	National Agency for Education (created in 1991) Swedish Schools Inspectorate (created in 2008)
United Kingdom (Northern Ireland)	Education and Training Inspectorate (part of Department of Education) Council for the Curriculum, Examinations and Assessment (created in 1993)

Source: Country Background Reports supplied by countries.

Some intermediate agencies have a large remit. For instance, in the Flemish Community of Belgium, the Agency for Quality Assurance in Education and Training (AKOV) is a central player in the evaluation and assessment framework. All services related to quality improvement of education fall under the jurisdiction of AKOV: determining and adjusting the attainment targets and developmental objectives, ensuring clear processes for certification, ensuring quality of educational institutions, organising the National Assessment Programme, ensuring coherence between the attainment targets and developmental objectives of compulsory schooling and competences of teachers taught in initial teacher education programmes, organisation of the Examination Board of the Flemish Community, outlining criteria for recognition of prior learning, and supporting the work of the Inspectorate Flemish (Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

An important issue for policy is the division of labour between education authorities and intermediate agencies such as quality or evaluation agencies or school inspectorates, for instance in terms of making a judgement on the state of education in countries and developing a vision for evaluation and assessment. Generally, while there is collaboration between education authorities and intermediate agencies, it is often ambiguous how far the latter can take their autonomy in leading educational evaluation activities. The reality is that education authorities inevitably have a vested interest in the evaluation of the education system which gives great importance to the issue of the independence of intermediate evaluation agencies.

The devolution of responsibilities for evaluation and assessment involves a variety of trade-offs

There is considerable variation among OECD countries in the extent to which the governance and implementation of evaluation and assessment are devolved to the local level. One implication of significant levels of devolution of responsibilities for evaluation and assessment are variations in the implementation of national policy for evaluation and assessment at the regional/local level. This has both advantages and drawbacks. The diversity of approaches to evaluation and assessment allows for local innovation and thereby system evolution and the large degree of autonomy given to the region, municipal and school levels may generate trust, commitment and professionalism. It might also encourage collaborative work within schools on the adaptation of evaluation and assessment procedures at the local level. At the same time, there may be concerns about the lack of systematic application of national directions, inconsistency of practices and little capacity or commitment to developing quality frameworks at the local level. These concerns might be amplified by weak articulations between the different decision-making levels (e.g. between regions and municipalities) and limited collaboration between the regions, municipalities and schools (e.g. as in networks and partnerships of municipalities to take collective responsibility for quality evaluation and improvement).

In highly decentralised countries, there are typically excellent quality assurance initiatives at the local and school level but a number of challenges arise. The school owners (e.g. regions, municipalities and private providers) generally vary in their capacity for and commitment to evaluation and assessment activities. The instruments used for quality assurance and reference standards might be extremely diverse. They are also often not documented at the national level. There might also be few mechanisms to identify good practice and share it across the entire system. In Hungary, school maintainers (local governments and independent maintainers) take responsibility for evaluating the effectiveness of the pedagogical work of schools and their professionals. However, no

criteria are defined at the national level and there is great autonomy to decide the specific approach followed. Public schools are required to follow the rules stipulated in the Local Council Quality Management Programme (Hungarian Ministry of Education and Culture, 2010).[1] In the Flemish Community of Belgium, the regulatory framework for quality assurance is developed to fit the Freedom of education principle, which grants the schools the right to develop their own concept of quality of education. This entails investments in the capacity of schools concerning school self-evaluations, use of data by school agents, and the development of school development plans (Shewbridge et al., 2011a).

Studies from the United States show the impact of school districts on school and student performance varies widely. Several studies point to the importance of district leadership in developing strategies for improvement, helping schools to align curriculum to central standards and assessments, and providing support for low-performing schools (see for example, Elmore and Burney, 1997; Hill et al., 2000; Newmann et al., 2001). Approaches at the school level might also be equally diverse. The ways in which school self-evaluation is conducted, the role that school leaders take in the framework and the communication channels between schools and school owners might be highly variable. At the classroom level, there might be some insecurity among teachers about how to best implement the curriculum and marking criteria so as to ensure a fair assessment of student performance in the case national student learning objectives leave considerable room for local interpretation.

Therefore, a major challenge for the evaluation and assessment framework lies in determining what constitutes a desirable measure of national consistency as against local diversity. National agendas for education are likely to be strengthened by greater consistency of evaluation and assessment procedures across schools but greater diversity offers more opportunities for innovation and adaptation to local needs. The challenge is to articulate a national strategy for each of the evaluation and assessment components which builds on the best of current practice and serves the national reform agenda and continues to allow flexibility of approach at the local level within agreed parameters. This involves ensuring strong willingness at the local level to build on the national evaluation and assessment agenda by adapting it to local needs and specificities.

Integration of the non-public sector

In most countries, a non-public sector co-exists in the provision of education services, often with subsidies provided by the public budget. Countries with significant private school provision include Australia (over 30% of enrolment in non-public schools, all of which receive some degree of government funding), Belgium (over 50%, publicly funded at the same level as public schools), Chile (over 50%, the majority of which are funded at the same level as public schools), Denmark (over 25% in lower secondary education, the large majority of which is government-dependent), France (over 30% in upper secondary education, the large majority of which is government-dependent), Japan (over 30% in upper secondary education, all of which are not publicly subsidised), Korea (over 46% in upper secondary education, all of which is government-dependent) and the United Kingdom (over 45% in upper secondary education, the large majority of which is government-dependent) (OECD, 2012a).

The integration of non-public schools in the overall evaluation and assessment framework varies considerably across countries. In Belgium, where the Freedom of education principle provides parents with free school choice, the evaluation and assessment framework treats all schools equally, i.e. all schools are subject to the same rules and

engage similarly in centrally dictated evaluation and assessment activities (in addition to the their own sector's specific initiatives). By contrast, in Chile, which places strong emphasis on parental freedom of school choice through its voucher programme, central requirements for evaluation and assessment in private schools are considerably more limited than in public schools. For instance, the approach for teacher appraisal in the private school sector (most of which is publicly subsidised) consists of giving independence to school providers to run their own procedures. However, private schools are required to participate in national standardised student assessments (Santiago et al., forthcoming).

In Denmark, most of the 2006 policy initiatives to strengthen the evaluation culture in compulsory education do not apply directly to private schools. While private schools have to demonstrate similar conditions for student assessment in as much as they must provide end objectives and educational descriptions of how students will reach these, evaluate the student's learning outcomes and communicate this with parents and evaluate the school as a whole on a regular basis, they have considerable freedom in how they do so. Private schools are not required to use the national student Common Objectives or national tests. Further, they may opt out of administering final examinations in Grade 9 by officially informing the Ministry of Education. The parents hold the primary responsibility for supervising the educational quality in private schools. Private schools choose between self-evaluation and a parent-elected certified supervisor. They are accountable to supervision by the Ministry of Education, specifically, the Quality and Supervision Agency (Shewbridge et al., 2011b).

In Australia, there is strong emphasis on working with all school sectors on all the key areas for schooling. Through the Schools Assistance Act 2008 non-government schools have an obligation to meet national school performance and reporting requirements similar to those which apply to government schools, including participation in national and international student standardised assessments, publication of school-level reports and reporting to parents. However, the integration of the non-government sector within the overall evaluation and assessment framework is more limited at other levels. For instance, school evaluation practices in the Catholic and Independent sectors may not be mandatory and the organisation of teacher appraisal in the context of performance management processes is dissociated from state and territory School Improvement Frameworks. The typical approach for teacher appraisal and school evaluation in the non-government sector consists of giving independence to school providers to run their own procedures while state and territory authorities monitor the performance of non-governmental schools against minimum standards through periodical registration processes and for their accreditation for credentialing students. The environment of choice under which non-government schools operate creates an imperative for continual evaluation and assessment in order to ensure individual schools continue to meet the needs and expectations of students, parents, the community and governments (Santiago et al., 2011).

The policy challenge is to ensure a degree of integration of non-public schools in the overall evaluation and assessment framework which guarantees that evaluation and assessment procedures in the non-public sector are sufficiently aligned with national student learning objectives and educational targets, while respecting the freedom of organisation of non-public schools and acknowledging that they are subject to market-based accountability to a greater extent. Such degree of integration into the overall evaluation and assessment framework is desirable as a way of providing information about schools which is comparable and, for non-public schools receiving public subsidies, as a way to hold schools accountable for the use of public funds.

Overarching reference: Goals for student learning

Clear and widely supported goals for student learning provide the solid reference point on which to build evaluation and assessment. They are expressed both at a more generic level (e.g. overall objectives of the education system) and in more specific ways (e.g. curricula). These are essential to achieve the alignment of processes and school agents' contributions within the evaluation and assessment framework.

General goals for the education system

Countries typically devise statements about the ultimate goals of their education system and governments generally establish priorities for education policy for the period they are in office. In addition, it is becoming increasingly common for governments to set up education targets alongside indicators to assess progress towards the targets.

Overall goals for education systems typically emphasise the following aspects:

- the personal development of individuals
- the acquisition of skills and competencies (e.g. learning in the course of life, critical thinking)
- equality of educational opportunities
- values and attitudes (e.g. civic skills, fundamental rights, principles of democracy, respect of diversity, protection of the environment).

Education policy priorities, often associated with specific education targets, generally address the following aspects:

- educational outcomes (e.g. completion rates, performance levels, quality of outcomes)
- equity of outcomes (e.g. outcomes for particular student groups)
- education processes (e.g. implementation of a reform; accountability and transparency; school leadership; quality of teaching)
- education staff (e.g. raising the status of teaching, working conditions)
- specific target areas (e.g. expansion of vocational education, strengthening of early childhood education).

For instance, in Mexico, the General Education Law provides clear goals for education emphasising the development of individuals and the promotion of values and attitudes. These are associated with broader social and economic goals. Statements about the aims for the education system such as its promotion of diversity, equity and quality and its role in developing successful learners and informed citizens are articulated. In addition, federal governments in office establish priorities for educational policy, which provide the framework for policy development. For instance, one of the six objectives of the 2007-2012 Education Sector Programme was "to promote the development and use of information and communication technologies in the education system in order to provide support for student learning, increase student life abilities and favour student entry into the knowledge society". Education targets to be achieved by 2012 were also established with associated indicators to permit the monitoring of their achievement (e.g. in primary

school, a proportion of 82% and 83% students achieving at the *basic* proficiency level in national tests, in Spanish and mathematics respectively) (SEP and INEE, 2011).

In Norway, the three core objectives for education, as defined by the Ministry of Education and research in its 2007-08 report to the Parliament on *Quality in Education*, refer to basic skills development, completion of upper secondary education and inclusion (Norwegian Directorate for Education and Training, 2011). In the Czech Republic, two of the seven objectives of the education system as defined by the 2005 Education Act are: "The personal development of a human being who shall possess knowledge and social competencies, ethical and spiritual values for their personal and civic life, for pursuing a profession or working activities, and for acquiring information and learning in the course of life"; and "The formation of national and state citizenship awareness and respect for the ethnic, national, cultural, language and religious identity of every person" (IIE, 2011). In Australia, the *Melbourne Declaration on Educational Goals for Young Australians* defines two overarching goals for schooling: "Australian schooling promotes equity and excellence"; and "All young Australians become successful learners, confident and creative individuals, and active and informed citizens" (DEEWR, 2010).

Equity goals in education systems

In most countries there is an emergent focus on equity and inclusion among educational goals. For example, in Australia, equity is at the core of the national goals for education, and national reporting on education pays careful attention to different measures of equity, including gender, Indigenous groups, geographic location, students with a language background other than English and socio-economic status (based on parental education and parental occupation). Equity has been given more prominence in general government reporting since 2004 when it was put on the same level as "efficiency" and "effectiveness" in the Report on Government Services' general performance indicator framework, with indicators on equity of access (output) and equity of outcomes (DEEWR, 2010). Also, in the Flemish Community of Belgium, there is strong political focus on the need to increase the equity of educational opportunities. Policy on Equal Educational Opportunities has played a prominent role since the adoption of the 2002 Decree of Equity of Educational Opportunities. The 2002 Decree provides for: the right for each child to enrol in the school of choice, with very strict rules on refusal or referral of students; the creation of local consultation platforms to ensure co-operation in implementing local equal educational opportunities policies; and extra support for schools providing additional educational support as part of this policy (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

However, in many countries, equity and inclusion are areas for further policy attention. In these countries, the articulation of equity among the national goals for education tends to be narrow. There is often limited knowledge about educational disadvantage in the education system – little differential analysis is undertaken on student performance across specific groups such as cultural minorities, students from disadvantaged families or those who live in a remote location. Also, in these countries typically no measures of equity in the education system have been developed so progress towards reducing inequities can be monitored.

Specific student learning objectives

At the level of student learning objectives, countries develop a basis for common expectations of outcomes from schooling, in a variety of forms such as curricula, study programmes, educational standards or learning progressions (see Chapter 4). For instance, in Mexico's primary and lower secondary education, there is a national curriculum supported by the general 2011 Study Plan for basic education and the grade- and subject-specific 2011 Study Programmes. These establish curricular standards to be met at the end of each of the four main stages of basic education as well as expected learning outcomes and are fairly detailed and prescriptive (Santiago et al., 2012b).

Other countries provide considerable more room for local adjustments to the curriculum. For instance, in the Slovak Republic, learning objectives for students are elaborated in common references established at the national level through national educational programmes. These binding documents consist of educational standards and stipulate the content of learning in each field of education and the expected outcomes at given stages in the education system. On the basis of the binding national educational programmes, schools then prepare more specific school educational programmes. These determine how the content proposed by national educational programmes is distributed into actual curricula for individual grades and subjects. Schools shape their profiles by means of their educational programmes (Shewbridge et al., forthcoming). In Italy, student learning objectives are defined in national curriculum goals, which can be adapted at the local level. Schools may devote up to 20% of their school time to the adjustment of the national curriculum to locally defined objectives, according to contextual factors which schools should analyse with relevant local stakeholders. For example, schools can now introduce an in-depth study of local historical events or to teach students certain skills to enter the local workforce.

Student learning objectives promote a range of competencies, which countries define in a variety of ways. For example, Austria defines "dynamic skills" which are transversal and not tied to specific subjects, Slovenia defines key competencies in thematic fields (e.g. learning to learn, social skills, ICT, entrepreneurship, environmental responsibility), Luxembourg defines foundation (*socle*) competencies as including both subject-based and cross-curricular competencies, and Finland has introduced the concept of "themes" (i.e. challenges with social significance) (Gordon et al., 2009).

As explained in Looney (2011a), several studies on "opportunity to learn" provide significant evidence that the focus, content coverage and flow, and cognitive demands in curricula have a strong and direct impact on student achievement (see Gamoran et al., 1997; Porter and Smithson, 2001; Smithson and Collares, 2007, cited in Schmidt and Maier, 2009). A priority is to ensure that student learning objectives are clear and detailed enough that the knowledge and skills students are expected to attain are readily apparent (Commission on Instructionally Supportive Assessment, 2001).

The development of student learning objectives is a complex exercise. For example, as explained in Looney (2011a), standards writers may have difficulty agreeing on the knowledge and skills that are most important. While the majority of OECD countries now promote skills for "learning-to-learn", including skills for problem-solving, critical analysis, as well as supporting students in developing greater autonomy, and so on, there may still be deep-seated tensions about the goals of education. Such "culture wars" (Finn and Kanstoroom, 2001), may lead to the development of standards that are vague (thereby avoiding controversy), or at the other extreme, standards that are overly detailed, making

it difficult to identify priorities for learning, and providing little useful guidance for instruction or the development of assessments (Chudowsky and Pellegrino, 2003).

There are also tensions between the idea of setting learning objectives for excellence for all students and supporting individual differences and interests. These are fundamental concerns for systems considering how to support both equity and quality (Linn, 1998). Policy makers may choose to set rigorous standards to communicate their efforts to raise school performance to the broader public. There is research supporting the view that students benefit from high expectations (Bransford et al., 1999). But there are also concerns that unreasonably high targets increase incentives for teachers to "teach to the test", thereby raising student scores, while not actually having an impact on student learning (Koretz, 2005) (discussed in more detail in Looney, 2011a).

Alignment between goals for student learning and evaluation and assessment

A crucial aspect for the successful implementation of evaluation and assessment is their alignment with student learning objectives. The core logic of criterion-referenced systems rests upon the alignment of goals for student learning, specific content for learning, pedagogical approaches and evaluation and assessment. The alignment is an effort of significant magnitude as it involves designing tools to accurately assess the competencies and expected learning outcomes promoted by national student learning objectives; fostering evaluation and assessment approaches consistent with the pedagogical approaches encouraged by learning goals (e.g. formative assessment); developing teacher capacity to assess against student learning objectives; designing instruments for teacher appraisal, school evaluation and school leader appraisal whose reference standards (typically teaching and school management standards, school evaluation frameworks) are aligned with student learning objectives; and ensuring all educational goals are covered in evaluation and assessment procedures (e.g. equity). The alignment might also involve the promotion of practices within the school system which support the achievement of educational goals such as a better use of evaluation results for feedback, greater focus on self-reflection by learners and educational practitioners, more interactive and collaborative work among school agents, and a closer focus on student competencies. Mexico introduced in the early 2010s a wide-ranging curricular reform with the potential to better align student learning objectives with pedagogical practices in schools and student assessment (see Box 3.1).

In a well-aligned system, student learning objectives are comprehensive, and clearly define the content and cognitive processes students are expected to demonstrate at different stages. If systems are poorly aligned, it is impossible for the system, schools and teachers to draw valid conclusions about student performance, or to adapt teaching to better meet identified needs (Linn, 2001, 2005; Haertel and Herman, 2005). Misalignment can have serious consequences on instruction and learning (Alliance for Excellent Education, 2010). For instance, one negative consequence of disconnected education standards and student standardised assessment is "teaching to the test", where teachers may emphasise assessment taking skills and low-level content, rather than important learning goals expressed by the standards (Alliance for Excellent Education, 2010). Also, if standards are poorly designed or not specific enough, teachers are more likely to focus on tests, thus narrowing the focus of teaching (Stecher et al., 2000).

Box 3.1 The Comprehensive Reform of Basic Education in Mexico

Mexico introduced in the early 2010s an extensive curricular reform to improve the coherence of the system and its focus on student achievement: the Comprehensive Reform of Basic Education (*Reforma Integral de la Educación Básica*, RIEB). Its key elements include the co-ordination among the different levels comprising basic education; the continuity between pre-primary, primary and lower secondary education; and the emphasis on issues relevant for today's society and education for life. The reform involves the preparation of updated study plans and programmes, focusing on pertinent teaching and with clearly defined expectations of skills to be acquired by year level and subject; improved training provided to school principal and teachers; and participative processes of school management.

The RIEB is based on a number of pedagogical principles, including: student-centred learning processes; planning to stimulate learning; creation of learning environments; collaborative work to develop learning; emphasis on the development of competencies and the achievement of curricular standards and expected learning outcomes; assessment for learning; promotion of inclusion to address diversity; reorientation of leadership; and pedagogical support to the school. The RIEB redefines learning as the development of competencies (instead of the transmission of knowledge contents), shifts pedagogical practices in classrooms and enhances reporting to students and parents. It puts emphasis on concepts such as assessment for learning, expected learning outcomes, collaborative learning, project-based work, student self-assessment and peer assessment and criterion-referenced marking, all of which place students at the centre of the learning. The RIEB is a profound structural educational reform, benefiting from the efforts of a large number of school agents, and drawing on the consensus achieved among educational stakeholders.

Source: Santiago et al. (2012b).

As explained in Looney (2009), no system can achieve perfect alignment. Baker (2004) points to the complexity of school systems as a barrier to tight alignment – including the number of links across different components of school systems, the diversity of regional contexts, differences in organisational contexts of schools, the range of teacher and school leader capabilities, resources devoted to professional development and other support for teachers. Given this complexity, it is very difficult to develop clear and explicit relationships across student learning objectives, instruction and assessments. Moreover, all assessments and examinations contain some degree of error, and as Haertel and Lorié (2004) assert, can only provide "an imperfect estimate of student performance". Another example of a source of misalignment is that educational measurement technologies have not kept pace with advances in the cognitive sciences, and large-scale assessments very often do not reflect educational standards that promote development of higher-order skills, such as problem-solving, reasoning and communication (Looney, 2011b). Misalignment, however, may in some circumstances establish a dynamic of change as when advances in the cognitive sciences lead to innovation in assessment (see also Chapter 4).

Alignment issues also apply to external school evaluation, school self-evaluation, teacher appraisal and school leader appraisal. References for school evaluation, school leader appraisal and teacher appraisal need to align with student learning objectives and include the associated criteria to be used by evaluators. Agencies in charge of school evaluation, school leader appraisal and teacher appraisal should also be able to provide evidence of inter-evaluator rating reliability. In the context of school self-evaluations, staff may need to achieve consensus regarding goals for the evaluation, and the criteria by which they will judge school performance. All these evaluation activities require good "social alignment", i.e. social capital in systems, including shared values, motives and efforts

around educational goals and the principles underlining them (Baker, 2004; Hargreaves, 2003). In socially aligned systems, institutions and actors work together to define challenges and to consider alternative courses of action. This alignment is vital for system learning and improvement (Looney, 2011a).

Design

Principles

In designing their evaluation and assessment frameworks, countries draw on a range of principles which are typically aligned with the overall goals for and traditions in their education system and are expected to improve the effectiveness of evaluation and assessment procedures. These include placing students at the centre of the evaluation and assessment framework, focussing on student outcomes, committing to transparency through the reporting of evaluation and assessment results, promoting a culture of sharing classroom practice, relying on teacher professionalism and responding to diverse learner needs.

Placing students at the centre of the evaluation and assessment framework

A desirable principle in the design of the evaluation and assessment framework is to place students at the centre of the evaluation and assessment framework. In Sweden, an important aspect of the approach to education is that students are being trusted and considered as responsible partners in the education system in general, and in evaluation and assessment activities in particular. The Education Act and the curriculum state that all students should be granted the democratic rights of taking responsibility and participating in the decisions that concern them and their school environment. Teachers are required to involve their students when planning and organising lessons. Even in the earliest grade levels, teachers discuss the goals and performance criteria with their students, the involvement of students being adapted to their age and maturity. Teachers are obliged to ensure that students and parents are well informed about the goals and receive regular feedback about their progress. Student assessment throughout compulsory education is organised around individual development plans (IDPs). These are developed and revised collaboratively in regular "development talks" between the teacher, the individual student and his or her parents. The goals determined in IDPs can also be used for student self-assessment in which students are asked to rate their own progress and performance. Students and their parents also play an important role in the evaluation of educational services. At the national level, the National Agency for Education carries out since 1993 regular surveys of student, parent and teacher attitudes towards school. The survey covers issues such as safety, comfort, atmosphere at school, teaching and learning, and opportunities for student participation. Locally designed student and parent surveys are also frequently used to gather the opinions and expectations of key client groups (Nusche et al., 2011a).

By contrast, in some other countries, teaching, learning and assessment still take place in a somewhat "traditional" setting with the teacher leading his/her classroom, the students typically not involved in the planning and organisation of lessons and assessment concentrating on summative scores. In such settings, while teachers may also focus on the development of skills for critical analysis and reasoning, students have few opportunities to experiment with or apply new knowledge in different situations. Knowledge is presented within traditional, discrete categories (language, mathematics, science, arts, and so on), and there may be few opportunities to link learning across subjects. There is little

attention paid to student motivation, or to developing skills for "learning to learn" (OECD, 2009). In these countries, the opportunity given to parents and students to influence student learning also tends to be more limited and there is little emphasis on the development of students' own capacity to regulate their learning. Other practices which might require further strengthening in these countries are the communication of learning expectations to students, the opportunities for performance feedback and mechanisms for individualised support. Also, collecting the views and perspectives of parents and students to inform school improvement through the systematic use of surveys is not a general practice in some countries. This includes surveys designed by teachers to collect student views on their teaching.

A practice which raises concerns of not placing the student at the centre of the learning is the extensive use of year repetition. According to PISA 2009 data, year repetition is particularly high in Belgium, France, Luxembourg, Portugal and Spain where over 35.0% of students had repeated one or more years by age 15 according to themselves (against an OECD average of 13.0%) (OECD, 2010). This raises important concerns. First, high levels of year repetition are not compatible with placing students at the centre of assessment as it extensively involves branding students a failure at different stages of schooling, including in the very early stages of learning. Second, it runs counter to the need for teachers to have the highest possible expectations of what children can achieve if they always have the possibility of retention in the back of their minds for children who do not respond well to their teaching. It should be recognised, however, that in those countries where levels of year repetition are high, there is typically a belief among teachers and parents that year repetition is beneficial for low-performing students, leading such practice to become part of the school culture.

There is wide recognition in educational research that year repetition is an ineffective intervention for low achievement while it poses risks for equity in terms of bias based on social background (Field et al., 2007; OECD, 2012b). Reviews of the research literature by Brophy (2006) and Xia and Kirby (2009) concluded the following about school-imposed year repetition:

- It improves academic achievement temporarily, but over time, year repeaters fall further and further behind other low achievers who were promoted.
- It is stressful to students and associated with reduced self-esteem, impairs peer relationships, increases alienation from school, and sharply increases likelihood of eventual dropout.
- It makes classes larger and harder to manage for teachers and creates budgetary and equity problems for schools and school systems.

Research in both the United States and France suggests that social background, independent of school attainment, is an important determinant of repeating. This may be due to behavioural difficulties associated with social background, or because educated parents are in a stronger position to oppose a repetition proposed by the school. Therefore year repetition may also pose risks for equity in terms of bias based on social background (Field et al., 2007; OECD, 2012b). Also, the costs of repetition for the education budget are substantial given the extra expenditure incurred in the repeated year and the opportunity costs of one year of the student's time. For example, in Italy, the Ministry of Education estimated that repetition in secondary school may result in a 6% increase of per student expenditure (MIUR, 2011). This is exacerbated by the fact that schools have very few incentives to take these large costs into account. In summary, year repetition is

ineffective and costly; this has both efficiency and equity implications (Field et al., 2007; OECD, 2012b).

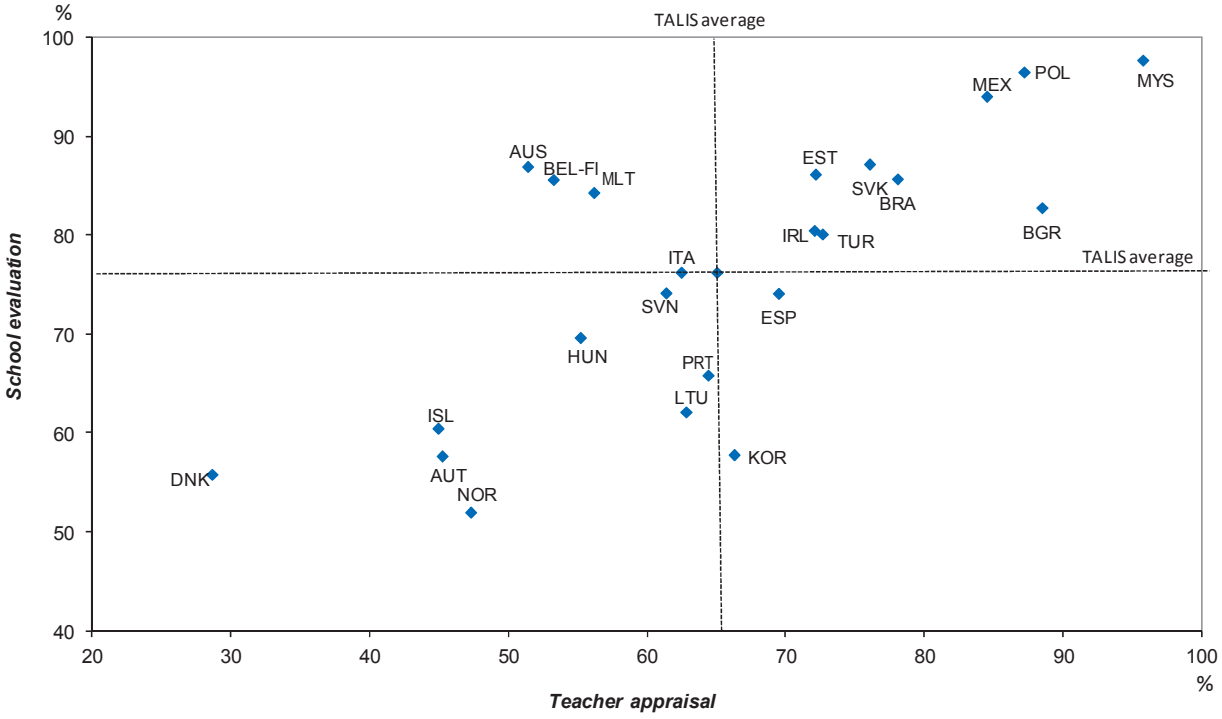
Focussing on student outcomes

Evaluation and assessment frameworks are increasingly focusing on improving student outcomes and achieving student learning objectives. This is reflected in the priorities for national monitoring (in particular the introduction of student standardised assessments), the establishment of educational targets, the significance of evidence on student performance for school evaluation and teacher appraisal, and the importance of reporting publicly on student results. Figure 3.2 provides an indication of the importance of student results in both teacher appraisal and school evaluation according to the perceptions of teachers and school principals in lower secondary education, for countries which participated in the 2007-08 OECD's Teaching and Learning International Survey (TALIS).[2] It suggests that the focus on student results is particularly important in OECD countries such as Mexico, Poland and the Slovak Republic and relatively less so in Austria, Denmark, Iceland and Norway.

Figure 3.2 Student test scores as a criterion of teacher appraisal and school evaluation (2007-08)

Vertical axis: Percentage of teachers of lower secondary education whose school principal reported that student test scores were considered with high or moderate importance in school self-evaluations or external school evaluations

Horizontal axis: Percentage of teachers of lower secondary education who reported that student test scores were considered with high or moderate importance in the appraisal and/or feedback they received



StatLink http://dx.doi.org/10.1787/888932791153

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Committing to transparency through the reporting of evaluation and assessment results

The overall evaluation and assessment framework can be strengthened through transparency in monitoring and publishing results. Reporting, as a key purpose of evaluation and assessment, is becoming increasingly important as reflected in requirements at several levels: system level (e.g. reports on the state of education, education databases); school level (websites with school-level information, school annual reports, inspection reports); and student level (publication of standardised assessment results at the school level, reporting of marks to students and parents).

Figure 3.3 displays the extent of publication of school evaluation results (including school performance tables) in countries which participated in the first round of TALIS in 2007-08, according to school principals' perceptions. It suggests that the emphasis on the public reporting of school evaluation results is particularly important in countries such as Australia, Belgium (Flemish Community), Denmark, Iceland, Mexico and the Slovak Republic. However, some countries emphasise the publication of student results at the school level (e.g. Mexico) while others emphasise the publication of qualitative reports (e.g. municipal quality reports in Denmark, inspection reports in the Flemish Community of Belgium). By contrast, reporting of school evaluation results is less common in countries such as Austria, Poland, Slovenia, Spain and Turkey. Figure 3.4 provides indications on the publication of student achievement data at the school level through the perceptions of school principals in PISA 2009. It reveals a clear contrast between countries with strong emphasis on the reporting of such data (e.g. Netherlands, New Zealand, United Kingdom, United States) and those where such practice is infrequent (e.g. Austria, Belgium, Finland, Japan, Switzerland).

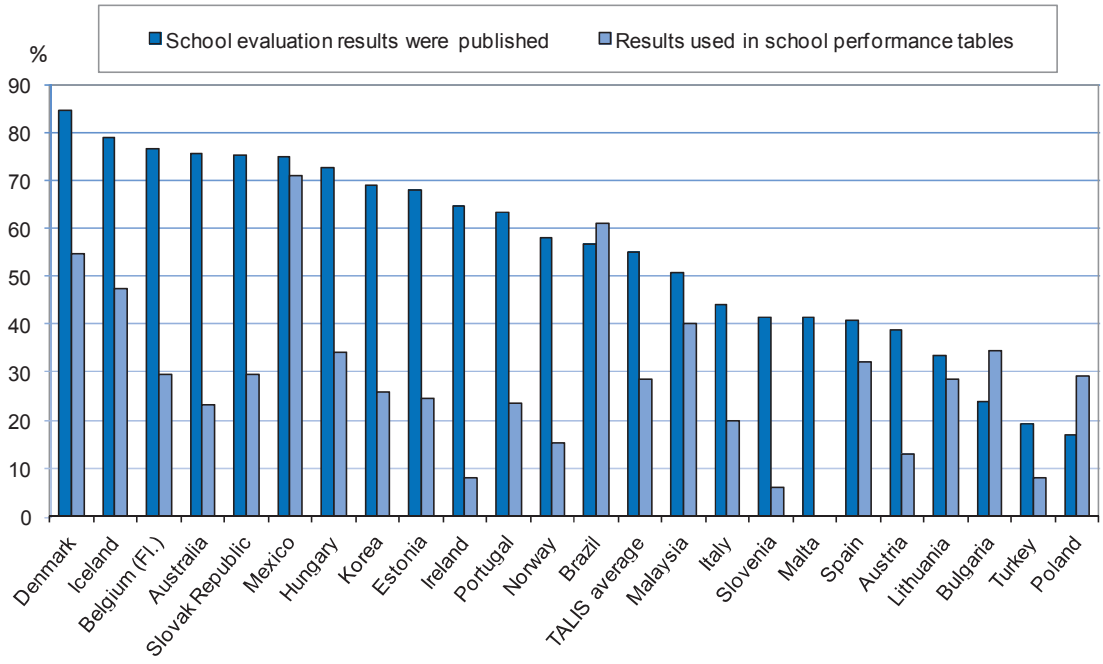
In Mexico, the evaluation and assessment framework is reinforced by the establishment of significant requirements for public reporting. Mexico collects a wide range of data on education system performance, including through international student surveys, national standardised assessments, qualitative studies and the development of educational indicators. Comprehensive sets of educational statistics are published and education databases have been developed. Furthermore, the results of national standardised assessments are published at the school level but with the drawback that the simple averages provided do not allow for the appropriate contextualisation of the results. This situation is in contrast to the period prior to 2000, largely characterised by the absence of public data on educational outcomes. From 2002 on, there has been an explicit objective of disseminating publicly data on educational outcomes both to hold school agents accountable and to ensure the respective analysis informs educational policy development (Santiago et al., 2012b).

In Australia, following the *Melbourne Declaration on Educational Goals for Young Australians* in December 2008 Ministers agreed that public reporting on Australian schools would: support improving performance and school outcomes; be both locally and nationally relevant; and be timely, consistent and comparable. In June 2009, they agreed a set of eight principles and related protocols for reporting on schooling in Australia, the *Principles and Protocols for Reporting on Schooling in Australia* (MCEECDYA, 2009). This is a document which makes clear their commitment to transparent accountability. The principles relate directly to data on student outcomes and information about the school context and resourcing. The protocols are designed to promote the integrity of the

process and to provide safeguards against simplistic comparisons being made amongst schools (Santiago et al., 2011).

Figure 3.3 Publication of school evaluations (2007-08)

Percentage of teachers of lower secondary education in schools where school evaluations were published or used in comparative tables



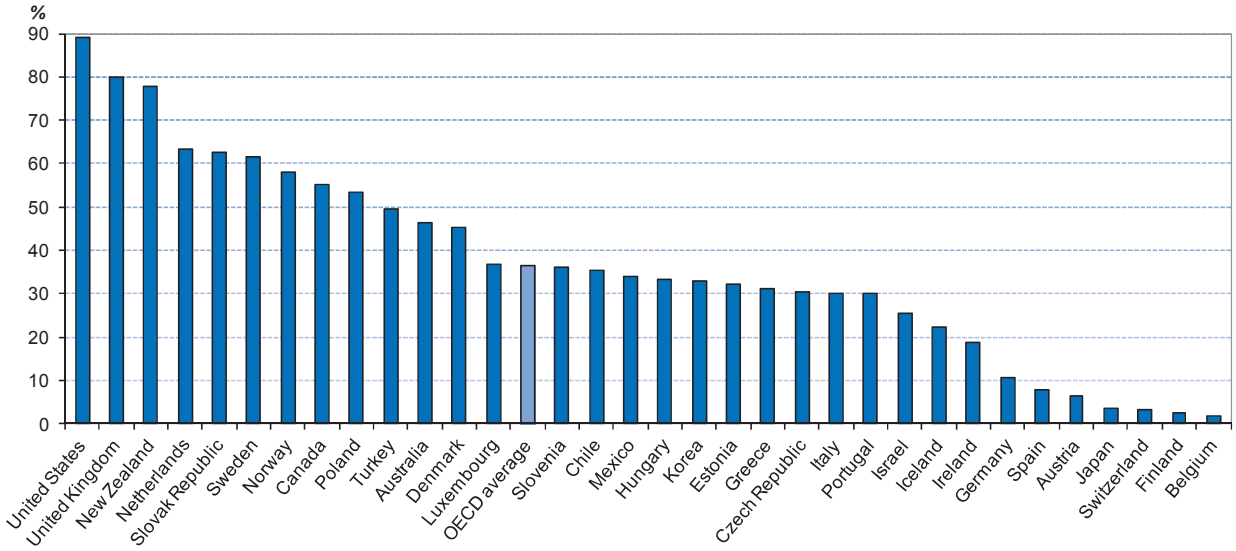
StatLink http://dx.doi.org/10.1787/888932791172

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

As explained by Rosenkvist (2010), it is a widely debated question in many countries to what extent and how student assessment results should be made publicly available. Some contend that there should be an effort towards making public all evidence from the evaluation of public policy (with appropriate analyses) in order to provide evidence to taxpayers and the users of schools on whether the schools are delivering the expected results, to enhance trust in government, or to improve the quality of the policy debate. Others consider that the publication of student results at the school level will be counterproductive as it is subject to erroneous interpretation, particularly when no adjustment for socioeconomic background is made, and may provoke some detrimental effects such as teaching to the test as a result of school agents concentrating on the measures which are published (see also Chapter 6).

Figure 3.4 Publication of student achievement data (2009)

Percentage of 15-year-old students in schools where the principal reported that student achievement data was posted publicly



StatLink http://dx.doi.org/10.1787/888932791191

Source: OECD, PISA Database, *Compendium for the School Questionnaire*, *http://pisa2009.acer.edu.au/downloads.php*.

Promoting a culture of sharing classroom practice

The evaluation and assessment framework benefits to a great extent on the ability to cultivate a culture of sharing classroom practice, developmental classroom observation, professional feedback, peer learning and professional coaching. Research shows that teachers are better able to adapt teaching to the needs of their students when they share information about instruction methods and student learning (Little, 1990; Newman and Wehlage, 1995; McLaughlin and Talbert, 2001). For example, the Consortium for Policy Research in Education in Chicago found that among low-performing schools that had been placed on probation, those that had previously developed strong cultures of peer collaboration were able to exit probationary status relatively rapidly (from 1996 to Spring of 1998). This requires an "open door" climate among teachers, critical to ensure that the evaluation of teaching and learning quality is central to evaluation and assessment frameworks. O'Day (2002) suggests that the structure and norms of many schools, where teachers work in "independent and isolated classrooms", buffers individuals and schools against change and prevents mutual learning. One of the most important findings from a major study conducted in the United States investigating a three-year pilot of a new teacher evaluation system in the Chicago Public Schools was that the most valuable part of the entire process, for both teachers and evaluators, was in the professional conversations that accompanied an observed lesson (Sartain et al., 2011).

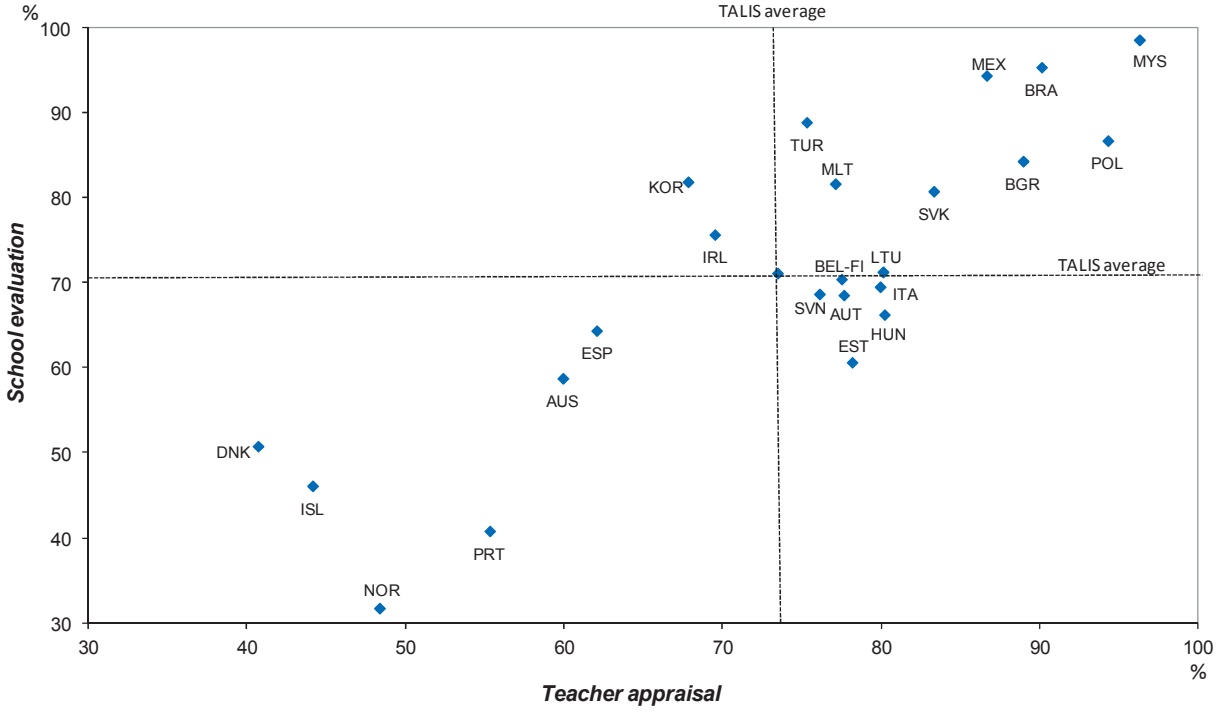
In some countries, classroom observations are not a systematic part of the appraisal of each teacher or are not conducted in the context of school evaluation and there is little tradition for school leaders or teacher peers to conduct informal classroom observations with an evaluative focus. Figure 3.5 displays perceptions by school agents on the use of

the direct appraisal of classroom teaching as a criterion of teacher appraisal and school evaluation, as part of the 2007-08 TALIS study in lower secondary schools. Countries in which there is less stress on the direct appraisal of classroom teaching include Australia, Denmark, Iceland, Norway, Portugal and Spain.

Figure 3.5 Direct appraisal of classroom teaching as a criterion of teacher appraisal and school evaluation (2007-08)

Vertical axis: Percentage of teachers of lower secondary education whose school principal reported that the direct appraisal of classroom teaching was considered with high or moderate importance in school self-evaluations or external school evaluations

Horizontal axis: Percentage of teachers of lower secondary education who reported that the direct appraisal of classroom teaching was considered with high or moderate importance in the appraisal and/or feedback they received



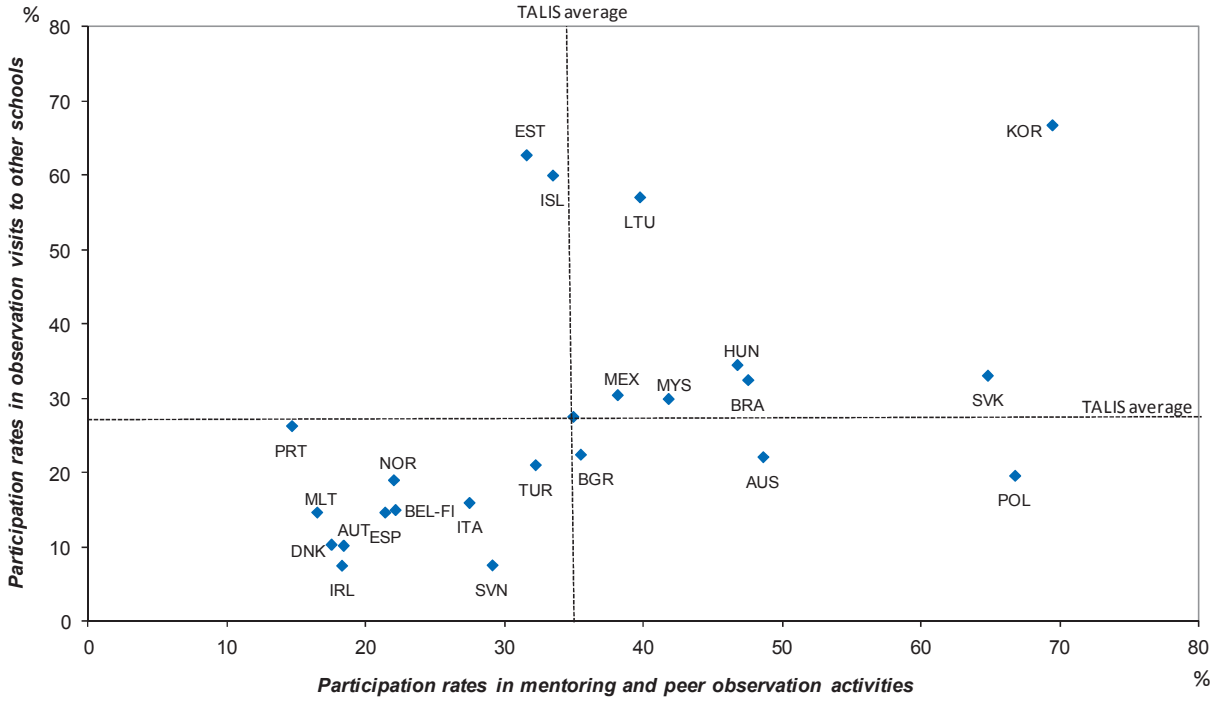
StatLink http://dx.doi.org/10.1787/888932791210

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Figure 3.6 provides another perspective on the sharing of classroom practice through the extent to which it is promoted by teachers' professional development activities, using TALIS data. It plots teachers' participation rates in mentoring and peer observation activities against their participation in observation visits to other schools. Teachers' feedback indicates that such activities might be significant in Hungary, Korea, Mexico and the Slovak Republic.

Figure 3.6 Participation rates in professional development activities involving the sharing of classroom practice (2007-08)

Percentage of teachers of lower secondary education undertaking specified professional development activities in the previous 18 months, international averages



StatLink http://dx.doi.org/10.1787/888932791229

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Relying on teacher professionalism

The ability for the evaluation and assessment framework to effect changes in the classroom and improve student learning depends to a great extent on its reliance on teacher professionalism. In New Zealand, the evaluation and assessment system is grounded in a strong belief in teacher professionalism. Teachers are seen as the main experts not only in teaching but also in assessing their students. This is in contrast to some other countries where student assessment is conceived to a great extent as an activity separate from teaching and undertaken by school-external psychometric experts. In New Zealand, the assessment of national standards is not based on whole cohort standardised assessment but instead relies on teachers' capacity to assess against the standards. National standards aim to provide external reference points of expected student performance while leaving the responsibility for choosing assessment methods and forming overall judgements with teachers. The approach to national monitoring (based on sample-based assessments) also involves teachers in the assessment activities. A range of teacher professional development programmes, as well as mentoring and induction for new teachers, aim to ensure strong teacher competencies in assessment. Teacher professionalism is further supported by well-established approaches to teacher appraisal. Teachers have a good degree of ownership of the appraisal process. It is the professional

body of teachers (New Zealand Teachers Council), and not an external agency that has taken the lead role in defining standards for teacher registration. Individual teachers are actively involved in their appraisal processes (both for registration and for performance management) through self-assessment of their own practices (Nusche et al., 2012).

In Sweden, the system of management by objectives requires strong teacher professionalism. Within the framework of the national steering documents and local plans, teachers have complete autonomy in deciding on teaching content, materials and methods. Moreover, teachers are seen as the main experts in assessing their students. For instance, while centrally developed national student tests exist in Sweden, they are administered and marked by the students' own teachers. Teachers are being trusted to review their own students' test performance and this is conceived as a way for them to further develop their pedagogical competencies. Teachers also play a key role in the internal evaluation of their own school. Quality assurance and reporting within schools have been conceived as a collective process with a strong focus on democratic participation and ownership by teachers (Nusche et al., 2011a).

A strong focus on professionalism implies the need for significant, sustained and focused investments in professional development. Teachers need to develop skills to assess learning needs and a broad repertoire of strategies to meet a range of student needs. Teacher professionalism also points to a stronger role for teachers in the development of student learning objectives and of assessment and evaluation systems. Based on their review of literature on accountability and classroom instruction, Ballard and Bates (2008) underscore the importance of communication among teachers and those who write learning objectives, develop large-scale assessments, and set out guidelines for school evaluations.

Responding to diverse learner needs

Evaluation and assessment have a key role in identifying and responding adequately to the needs of all students in the education system. Particular attention is typically given to groups for which underperformance is identified, students from cultural or language minorities, and students with special educational needs (see also Chapter 4). For example, standardised student assessment has the potential to disadvantage certain groups of students by generating a bias which is associated with a particular characteristic of the student such as gender, ethnicity, physical disabilities, and language of instruction differing from primary language. Le and Klein (2002) state that a fair student testing system accounts for three conditions: *(i)* test items are free of bias; *(ii)* students must have equal opportunities to demonstrate skills; *(iii)* students must have "sufficient opportunity" to learn the tested material.

In Ireland, the identification of students with special educational needs is a key aspect of the assessment framework for primary and secondary education (Irish Department of Education and Skills, 2012). In New Zealand, much work has been undertaken to develop assessment tools and approaches that are adapted to different learner groups. For example, the Ministry of Education is working with Māori assessment experts to develop approaches for monitoring student outcomes in the context of the curriculum and the standards used in Māori-medium settings. There is also a focus on developing guidance and resources for teachers to develop narrative assessment approaches and Individual Education Programmes for students with special educational needs. A number of language and literacy assessment tools are also available to provide adequate assessment opportunities for English language learners. There has also been some focus on including attention to Māori learner needs in teacher standards and teacher appraisal procedures.

The Registered Teacher Criteria emphasise the bicultural context of New Zealand. School review practices have also been adapted to ensure that school reviews fulfil the commitment of the education sector to improving education outcomes for Māori and Pasifika students. System evaluation focuses attention on ensuring that information is collected not only on the whole group of students but also on specific groups, and in particular the Māori and Pasifika students. This is intended to provide relevant information to identify strategies to respond to diverse learning needs (Nusche et al., 2012).

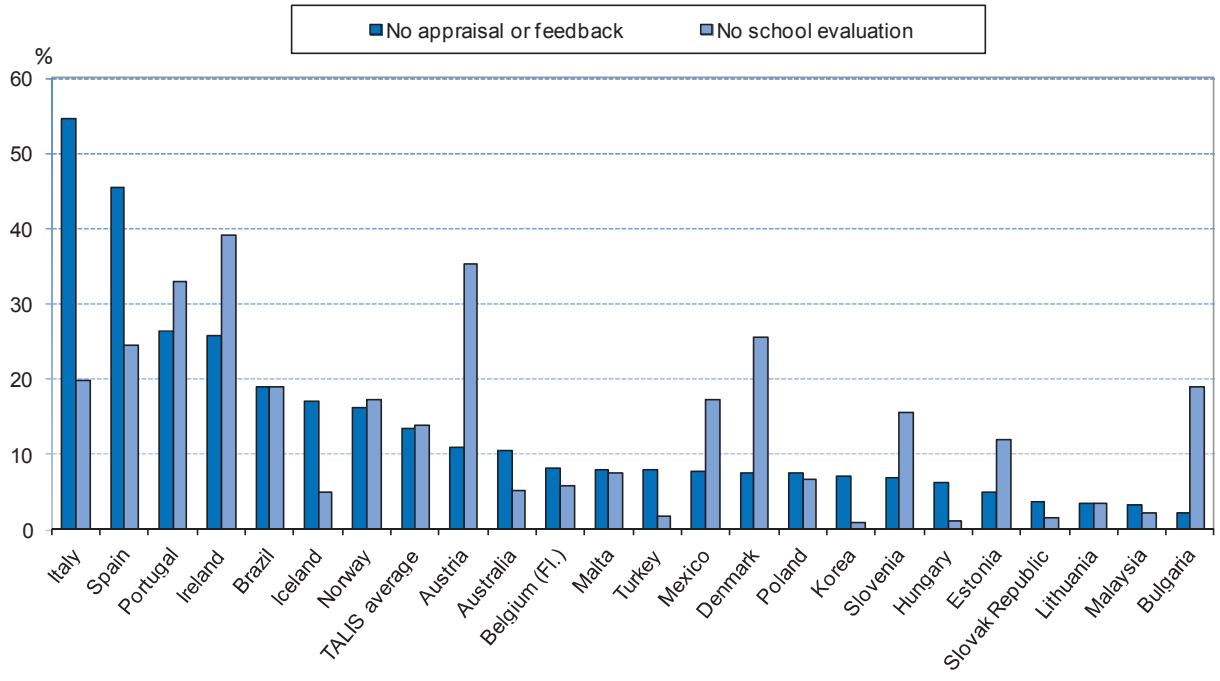
Main components

Typical components of the overall framework for evaluation and assessment are:

- **Student assessment**. Student assessment in general includes national (full-cohort or sample-based) standardised assessments, with diagnostic and monitoring purposes, and externally based summative assessment, in particular in view of assessing students for secondary education certification. At the school level, student assessment plays the key role in informing schools and teachers about students' individual achievement through teacher-based summative and formative assessments.
- **Teacher appraisal**. Procedures vary considerably across countries but, in addition to probationary processes, typically occur in two specific instances: (1) as part of performance management processes, including regular appraisal, to gain and maintain registration/accreditation to teach, and for promotion; and (2) to identify a select number of high-performing teachers to reward and acknowledge their teaching competence and performance. These formal schemes are often complemented with more informal school-level practices of feedback to teachers.
- **School evaluation**. There are generally two main forms of evaluation: school self-evaluation and school external evaluation or review. The latter is typically conducted by an external agency and involves a sequence of activities which may begin with self-reflection by the school, includes a visit by an external evaluator or team of evaluators, and leads to a summative report which may be published and may require a follow-up process. Schools may also be held accountable on the basis of comparable measures of student results.
- **The appraisal of school leaders**. Procedures vary considerably across countries but, in addition to probationary processes, are typically part of the employer's performance management processes with emphasis on administrative and pedagogical leadership.
- **System evaluation, the evaluation of sub-national education systems, and the evaluation of a programme or a policy**. In general, education system evaluation involves: *(i)* the monitoring of student outcomes at a given point in time, including differences among different regions within the education system and given student groups (e.g. by gender, socio-economic or immigrant background); *(ii)* the monitoring of changes in student outcomes over time; *(iii)* the monitoring of the impact of given policy initiatives or educational programmes; *(iv)* the monitoring of demographic, administrative and contextual data which are useful to explain the outcomes of the education system; *(v)* the development of means through which the relevant information is provided to the different agents in the education system; and *(vi)* the use of the generated information for analysis, development and implementation of policies.

A challenge in some countries is that some key components of a comprehensive evaluation and assessment framework are underdeveloped or do not exist. For example, the formative assessment of students by teachers might be underdeveloped as a result of the focus placed on results and a classroom practice dominated by examination and test preparation. Teacher appraisal might also not be systematic, school self-evaluation might be undertaken with little capacity, and the appraisal of school leaders might be at an incipient stage of development. Figure 3.7 provides an indication of the extent of teacher appraisal and school evaluation procedures in countries which participated in the 2007-08 round of TALIS according to the perceptions of teachers and school principals. It suggests that in some countries such as Ireland, Italy, Portugal and Spain, teacher appraisal and school evaluation were not systematic practices across schools at the time the TALIS survey was conducted. The situation may have changed in the case of these countries since then. For example, post-primary schools in Ireland are now regularly evaluated using a range of inspection models and Portugal is in the process of implementing a teacher appraisal model.

Figure 3.7 Teachers who received no appraisal or feedback and teachers in schools that had no school evaluation in the previous five years (2007-08)



StatLink http://dx.doi.org/10.1787/888932791248

Source: Reproduced from OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Table 3.3 provides a taxonomy of country approaches in terms of the comprehensiveness and degree of structure of their evaluation and assessment frameworks. The taxonomy uses information supplied by countries on features of evaluation and assessment frameworks (as displayed in Chapters 4 to 8). The assessment depends on the degree of existence of a national/state framework for the key components of an evaluation and assessment framework (*comprehensiveness*) and the extent to which evaluation and assessment practices are consistent across school agents and schools for the existing components (*degree of structure*). Some countries appear to have comprehensive and highly structured evaluation and assessment frameworks (e.g. Australia, Canada, France, Korea, Israel, New Zealand) while others opt for less comprehensive and less structured approaches (e.g. Austria, Italy).

Table 3.3 Comprehensiveness and degree of structure of evaluation and assessment frameworks across countries (2012)

		Comprehensiveness		
		High	Moderate	Low
Degree of structure	High	Australia, Canada, France, Korea, Israel, New Zealand	Chile, Estonia, Ireland, Mexico, Sweden	Luxembourg
	Moderate	Netherlands, Portugal, Poland, Slovenia	Belgium (Fl.), Belgium (Fr.), Norway	Denmark, Finland, Iceland, Spain
	Low	Slovak Republic	Czech Republic, Hungary	Austria, Italy

Definitions: A country's evaluation and assessment framework is considered of *high, moderate* or *low* comprehensiveness depending on the extent to which national/state frameworks exist for the key components of an evaluation and assessment framework. This involved the computation of an index on the basis of the information supplied by countries on features of evaluation and assessment frameworks and information from *Education at a Glance 2011*. The index considered the existence of national/state frameworks for student assessment (internal summative assessment, formative assessment, reporting of summative results, standardised central examinations, national student assessment), teacher appraisal (performance management and rewards, registration, probation), school evaluation (self-evaluation, external evaluation), appraisal of school leaders (performance management and rewards) and system evaluation (strategic collection of information, stakeholder surveys, national reporting).

A country's evaluation and assessment framework is considered to have a *high, moderate* or *low* degree of structure depending on the extent to which, for the existing components of the evaluation and assessment framework, practices are consistent across school agents and schools. This involved the computation of an index on the basis of the information supplied by countries on features of evaluation and assessment frameworks and information from *Education at a Glance 2011*. The index considered the degree of consistency of practices in student assessment (use of reference standards, moderation of marks, standardised reporting practices), teacher appraisal (existence of reference standards, consistency of aspects appraised and instruments, and identification of evaluators), school evaluation (extent to which self-evaluation and external evaluation are structured) and school leader appraisal (same criteria as for teacher appraisal).

Source: Taxonomy developed using information supplied by countries on features of evaluation and assessment frameworks (synthesised in Chapters 4 to 8) and information from *Education at a Glance* (OECD, 2011).

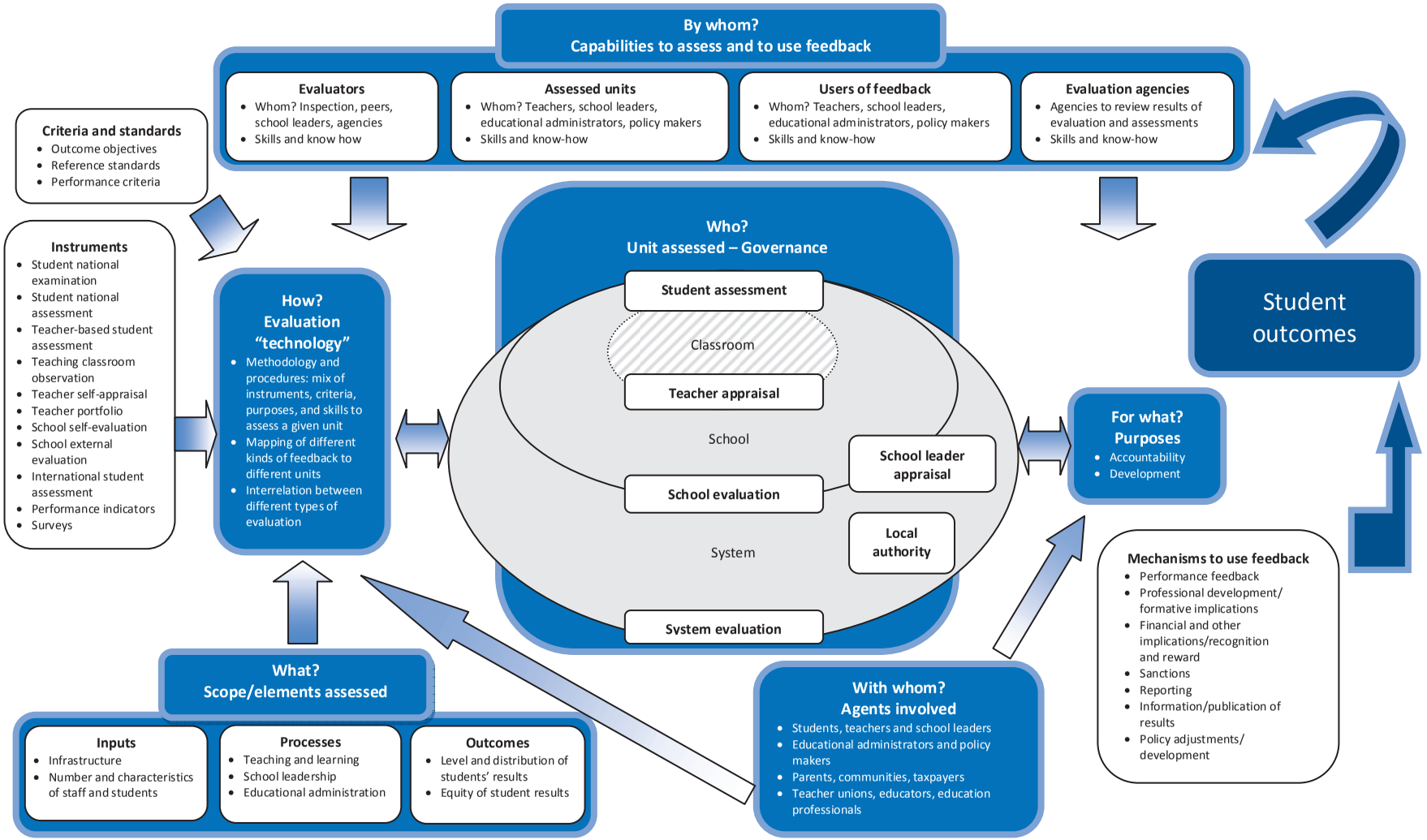
Main elements within evaluation and assessment components

Evaluation and assessment components have similar structures with key common constituents. Figure 3.8 summarises the main elements which are part of an evaluation and assessment component within the context of the overall evaluation and assessment framework. These are:

- ***Governance***: *Who?* This aspect concerns the overall design and organisation of one particular component of the evaluation and assessment framework. This includes the setting of the respective requirements and the distribution of responsibilities for the design of such evaluation component. It also includes the strategic goals of the particular evaluation component, its contribution to the overall evaluation and assessment framework and the balance between developmental and accountability functions.
- ***Capabilities to assess and to use feedback***: *By Whom?* This aspect concerns the preparation to evaluate, to be evaluated and to use the results of an evaluation as well as the choice of the groups undertaking these functions. It includes issues such as: the choice of the evaluators and the development of the skills to perform an assessment; the preparation to be the subject of an evaluation; the development of competencies to effectively use the results of an evaluation for the improvement of practice; and the design of agencies to review evaluation results with a view to hold agents accountable and to inform policy development.
- ***Scope***: *What?* Some evaluation processes concentrate on analyses of outcomes such as the level and distribution of students' results. Other evaluation initiatives centre on processes such as the quality of teaching and learning or the effectiveness of school leadership.
- ***Evaluation "technology"***: *How?* This aspect refers to the organisation of particular types of evaluation, that is the mix of instruments, criteria and standards, purposes, skills, and scope which are used to undertake a given evaluation or assessment. For instance, a teacher appraisal model may be based on a range of instruments such as self-appraisal, classroom observation and a teacher portfolio; be focussed on the teaching and learning process; be undertaken in relation to reference standards for the teaching profession; have both development and accountability purposes; and be based on experienced peers. Hence, this aspect refers to the way different aspects are combined to produce a given evaluation model.
- ***Purposes***: *For what?* This encompasses the objectives of a particular evaluation process and the mechanisms designed to ensure that evaluation results are used in a way such objectives are reached. The objectives of an evaluation process typically consist of development and accountability. Examples of mechanisms to use evaluation results include performance feedback, professional development plans, financial and other rewards, publication of results to the general public, and policy adjustments.
- ***Agents involved***: *With whom*? This mostly deals with the political economy of reform aspects of evaluation and assessment procedures. It relates to the involvement of a range of stakeholders such as parents, students, teachers, school leaders, teacher unions, educational administrators and policy makers in the development and implementation of evaluation and assessment processes.

This conceptual framework is used in subsequent chapters to analyse each of the components of the evaluation and assessment framework.

Figure 3.8 Conceptual framework to analyse evaluation and assessment in school systems

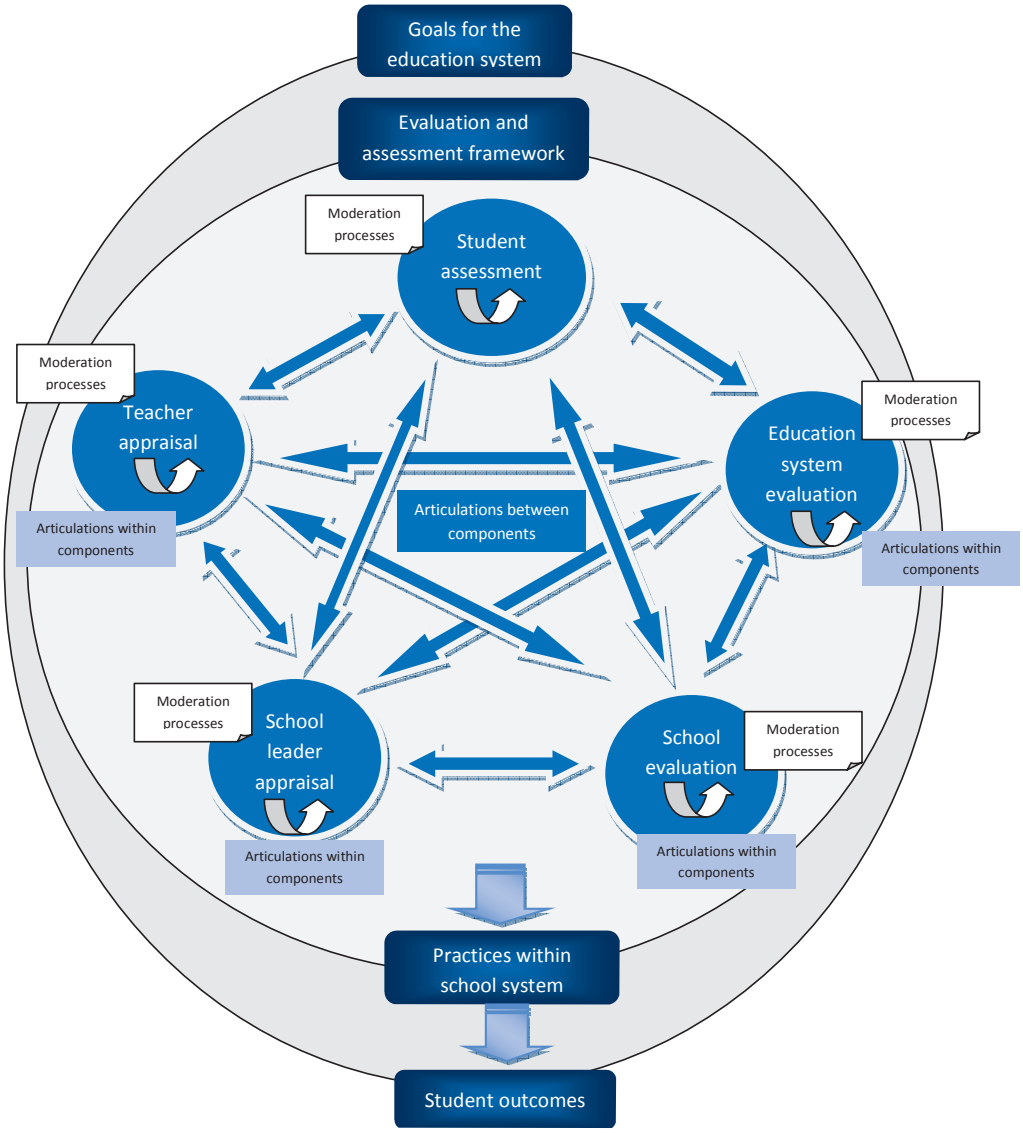


Articulations

How the different components have to be interrelated in order to generate complementarities, avoid duplication, and prevent inconsistency of objectives is an important aspect of designing the evaluation and assessment framework. Every country has some provisions for student assessment, teacher appraisal, school evaluation, school leader appraisal and system evaluation, but often these are not explicitly integrated and there is no strategy to ensure that the different components of the framework can mutually reinforce each other. A strategic approach to the development of the evaluation and assessment framework should provide an opportunity to reflect on the articulations between different evaluation components. Policy development needs to involve a reflection on ways to articulate the evaluation and assessment components to achieve the purposes of the framework.

Links and articulations within the evaluation and assessment framework can be categorised in three types: *(i)* articulations within components; *(ii)* articulations between components; and *(iii)* moderation processes to ensure the consistency of application of procedures within each evaluation and assessment component. This is portrayed in Figure 3.9.

Figure 3.9 Articulations within the evaluation and assessment framework



Articulations within components

The effectiveness of each component within the evaluation and assessment framework requires the establishment of linkages between its main constituents. Table 3.4 illustrates key articulations within components of the evaluation and assessment framework. A crucial aspect is to ensure each component is adequately aligned to education goals and student learning objectives.

Table 3.4 Key articulations within components of the evaluation and assessment framework

Student assessment	– Alignment between student learning objectives and student assessment – Linkages between student summative assessment and student formative assessment – Linkages between classroom-based assessment and external assessment – Alignment between performance ratings (marks) and educational standards – Alignment between student assessment and assessment courses in teacher education
Teacher appraisal	– Alignment between teaching standards and student learning objectives – Alignment between teaching standards and teacher appraisal – Systematic linkages between teacher appraisal and professional development – Alignment between teaching standards, registration processes and career structure – Articulation between school-based teacher appraisal and externally driven teacher appraisal – Linkages between formative teacher appraisal and *high-stakes* teacher appraisal – Alignment between skills taught in teacher education and teaching standards assessed in teacher appraisal
School evaluation	– Alignment between nationally agreed criteria for school quality and student learning objectives – Alignment between nationally agreed criteria for school quality and school evaluation – Articulation between school self-evaluation and external school evaluation – Systematic linkages between school evaluation and school development – Systematic linkages between school evaluation and school reporting – Complementarity between school performance measures and school review or inspection – Articulation between distinct school evaluation processes conducted by education authorities at different levels
School leader appraisal	– Alignment between school leadership standards and student learning objectives – Alignment between school leadership standards and school leader appraisal – Systematic linkages between school leader appraisal and professional development – Alignment between school leadership standards, registration processes and career structure – Linkages between formative school leader appraisal and *high-stakes* school leader appraisal
Education system evaluation	– Alignment between education goals, student learning objectives and education system evaluation – Placing education system evaluation in the broader context of public services evaluation – Mapping between education goals and targets and measures of education system evaluation – Systematic linkages between education system evaluation and education policy development – Systematic linkages between education system evaluation, public reporting and information systems – Coordination between the evaluation of the education system and the evaluation of education sub-systems – Complementarity between quantitative measures and qualitative system evaluation – Articulation between full-cohort and sample-based standardised assessments for national monitoring

An example of articulation within an evaluation component is the complementarity between school self-evaluation and school external evaluation. If the two processes are developed separately, this may lead to a costly duplication of data gathering and evaluation processes and the potential for external evaluation to test, affirm, strengthen and broaden school self-evaluation is not realised. In addition, the role that external evaluation can play in helping schools build their evaluative capacity and report progress effectively is also limited.

Indeed, external evaluation can potentially play a key role in reinforcing and supporting school self-evaluation by either validating or challenging the school's own findings. As a result, there is a need to reflect about the relative contributions of self-evaluation and external evaluation, ensure both processes use a common "language" (criteria for school quality), and define the nature of externality for school evaluation. Further linkages include ensuring that external evaluation includes a specific domain for analysis dedicated to "self-evaluation and improvement" and that external evaluation considers the school's own assessment of strengths and weaknesses.

Another example refers to the alignment of teaching standards with teaching career structures to reinforce the links between teacher appraisal, professional development and career development. This translates into an articulation between the definition of skills and competencies at different stages of the career (as reflected in teaching standards) and the roles and responsibilities of teachers in schools (as reflected in career structures) providing a clear structure for teacher appraisal.

Articulations between components

Synergies within the evaluation and assessment framework can also be realised through linkages between components. Table 3.5 illustrates key articulations between components of the evaluation and assessment framework.

An example of articulation across evaluation components is the mutually reinforcing linkage between school evaluation and teacher appraisal. This relates to a range of aspects such as: school-based teacher appraisal being validated by school evaluation processes; making the focus of school evaluation on teacher effectiveness systematic across schools; and school development processes exploring links to the evaluation of teaching practice. Another example is the articulation between school evaluation and the appraisal of school leaders. The results of school evaluation can usefully inform the appraisal of school leaders given that it reviews the performance of the school led by the leader being appraised and typically includes the assessment of leadership.

Table 3.5 Key articulations between components of the evaluation and assessment framework

	Student assessment
Teacher appraisal	– Student assessment results to inform teacher appraisal – Teacher's skills for student assessment to be systematically reviewed in teacher appraisal – Teacher's skills to use student assessment data to be reviewed in teacher appraisal – Teacher appraisal results to inform professional development on competencies for student assessment
School evaluation	– School evaluation to review student assessment practices in schools, including school-based student assessment criteria and teachers' skills for student assessment – School evaluation to review school's ability to moderate marking within the school and involve teachers in collaborative work around student assessment – School evaluation to review school's capacity to analyse and use student assessment data – Student assessment results to inform school evaluation – School reporting to include student assessment results – School evaluation results to inform school capacity development on competencies for student assessment
School leader appraisal	– Student assessment results to inform school leader appraisal – School leader's skills for implementing student assessment practices in the school to be systematically reviewed in school leader appraisal – School leader appraisal results to inform professional development on skills for evaluation and assessment
Education system evaluation	– Student assessment as part of the national monitoring of student outcomes, namely through standardised national student assessment – Educational trajectories of students as part of system evaluation (longitudinal studies) – Secondary analyses of student assessment data, including differentiated analyses across student subgroups – Results of education system evaluation to inform both the approaches to the national monitoring of student outcomes and the competencies to be assessed – Standardised national student assessment to assist the moderation of classroom-based student assessment – Policy evaluation to assess coherence of classroom-based student assessment across schools and its alignment with student learning objectives

	Teacher appraisal
School evaluation	– Validation of school-based teacher appraisal by external school evaluation – Review of school's application of external teacher appraisal by school evaluation processes – Focus of school evaluation on teacher effectiveness to be systematic across schools – School development processes, including school self-evaluation to use results of teacher appraisal – Results of school evaluation to inform the development of school-based teacher appraisal – Results of teacher appraisal to inform external school evaluation
School leader appraisal	– Teacher appraisal and school leader appraisal frameworks to inform each other – School leader appraisal to assess role of school leader in teacher appraisal, including the capacity to provide individual professional feedback and to lead the school's strategies to improve teacher effectiveness – Results of teacher appraisal to inform school leader appraisal
Education system evaluation	– Evaluation at the system and sub-system levels to use the information generated by teacher appraisal – Policy evaluation to assess consistency of teacher appraisal across schools – Results of education system evaluation to inform both the approaches to teacher appraisal and the competencies to be assessed by teacher appraisal

Table 3.5 Key articulations between components of the evaluation and assessment framework (*continued*)

	School evaluation
School leader appraisal	– The results of school evaluation can usefully inform the appraisal of school leaders – School evaluation to assess school leadership in schools – School leader appraisal to assess role of school leader in school evaluation, including pedagogical leadership – Sharing of aspects and criteria used to describe effective school leadership
Education system evaluation	– Evaluation at the system and sub-system levels to use the information generated by school evaluation – School annual reports to be used for system monitoring – Education system evaluation to include contextualised analysis of performance across schools, including with the identification of the factors shaping performance – Instruments such as national standardised student assessment providing the basis for fair comparisons across schools – Results of education system evaluation to inform both the approaches to school evaluation and the school quality criteria used in school evaluation

	School leader appraisal
Education system evaluation	– Evaluation at the system and sub-system levels to use the information generated by school leader appraisal – Policy evaluation to assess consistency of school leader appraisal across schools – Results of education system evaluation to inform both the approaches to school leader appraisal and the competencies to be assessed by school leader appraisal

Moderation processes

Another key aspect of designing an evaluation and assessment framework consists of the establishment of moderation processes to ensure the consistency of application of evaluation and assessment processes across schools and educational jurisdictions. In the implementation of teacher or school leader appraisal, it is important to ensure that appraisals against teaching or school leadership standards are consistent across schools and jurisdictions. This relates to the extent of externality in teacher and school leader appraisal and local interpretations of common standards with risks of lack of coherence of judgements. Similarly, ensuring the consistency of teacher-based student assessment within and across schools requires the establishment of moderation processes and the development of teacher capacity to assess against education standards (see also Chapter 4). These concerns also apply to external school evaluations, requiring school evaluation agencies to establish measures to ensure consistency of judgements against school evaluation criteria, and to system evaluation when applied to different educational jurisdictions.

Linkages to the classroom

Evaluation and assessment frameworks have no value if they do not lead to the improvement of classroom practice and student learning and therefore securing effective links to classroom practice is one of the most critical points in designing the evaluation and assessment framework. Examples of potential channels through which the evaluation and assessment framework impacts on classroom practice are assuring schools engage in meaningful self-evaluation practices and building teacher capacity for student formative assessment.

In some countries, the focus has been on structures, procedures, programmes and resources defined at the central level but, while these components of policy are clearly important, there has been a less clear articulation of ways for the national agenda for education to generate improvements in classroom practice through the assessment and

evaluation procedures which are closer to the place of learning. A strong top-down national vision for evaluation and assessment might constrain the ownership of evaluation and assessment procedures by school agents and result in a greater focus on the accountability function. This might also reflect the greater technical capacity at the centre and the more limited competencies at the local level to engage in evaluation and assessment activities. However, establishing links between evaluation and assessment and classroom learning inevitably requires establishing clear roles for local structures – school management, school supervision, local education authorities – in the implementation of evaluation and assessment policies. The point is that the fulfilment of the developmental function of evaluation and assessment requires articulation at the local level.

Capacity for evaluation and assessment

Competencies for evaluation and assessment

The effectiveness of evaluation and assessment relies to a great extent on ensuring that both those who design and undertake evaluation activities as well as those who use their results are in possession of the proper skills and competencies. This is crucial to provide the necessary legitimacy to those responsible for evaluation and assessment. Since evaluation has strong stakes for the units assessed and since school outcomes heavily depend on individual relations and co-operation at the school level, successful evaluation and assessment procedures require particular attention to developing competencies and defining responsibilities in evaluation processes. In addition, competencies for using feedback to improve practice are also vital to ensure that evaluation and assessment procedures are effective.

In most countries, while there have been efforts to strengthen assessment and evaluation activities, as well as providing competency-building learning opportunities in some cases, there are still limited evaluation and assessment competencies throughout education systems. Capacity building needs for evaluation and assessment are extensive and cover a range of areas such as:

- teacher capacity to assess against the whole range of curriculum goals to ensure consistency of marking across schools
- teacher capacity for formative assessment
- data handling skills of school agents (e.g. use of results from student standardised assessments)
- information for parents and other stakeholders to gain a good understanding of some outcome reporting
- capacity for taking on the role of external evaluator (e.g. in school evaluation, teacher appraisal)
- evaluation competencies of groups or agencies undertaking evaluation activities such as school evaluation or teacher appraisal, including school governing boards
- standardised assessment development, educational measurement, psychometrics, validation of test items, scaling methods
- externally based assessment (e.g. national examinations for certification)
- analytical capacity for educational planning and policy development.

As an example, Figure 3.10 provides teachers' perceptions of professional development needs in student assessment practices based on TALIS data for 2007-08. In OECD countries such as Italy, Korea, Norway and Slovenia, over 20% of teachers indicate that they have a high level of need for professional development in student assessment practices.

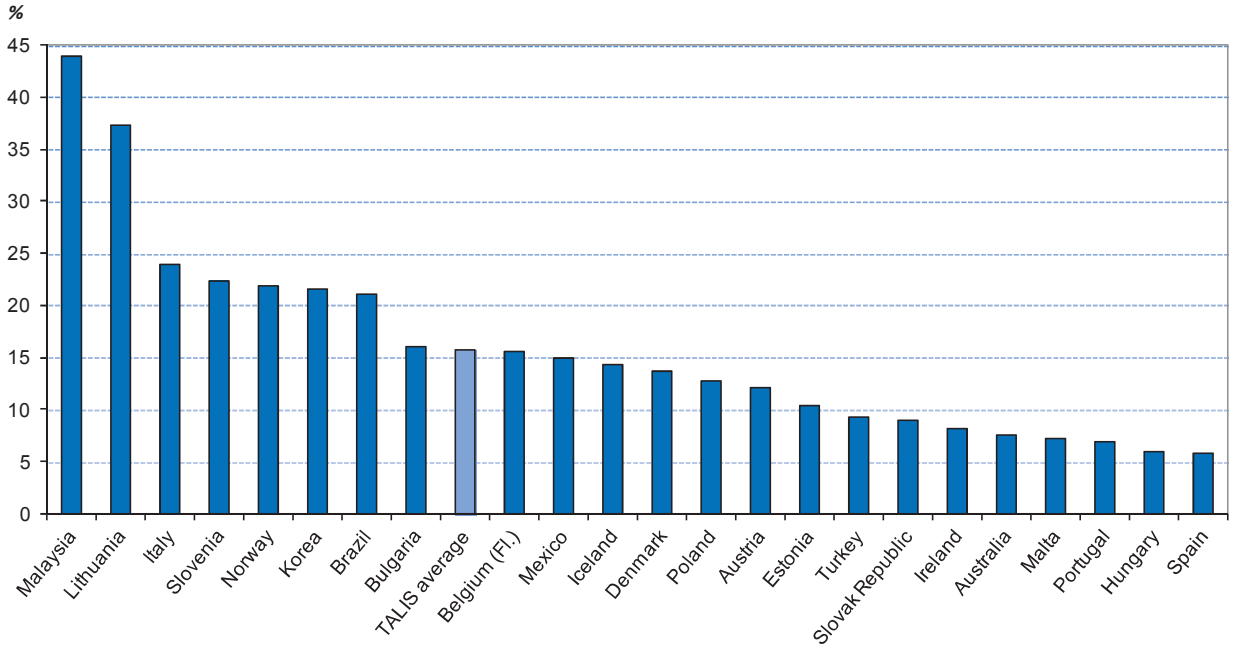
An area in which there is a growing need to develop school leader and teacher capacity is the ability to interpret and use data from standards-based assessments. Diagnosing the source of student difficulties and developing appropriate remedies for different students is often challenging. The process of developing "assessment literacy" typically encompasses the following actions: capacity to examine student data and make sense of it; ability to make changes in teaching and school practices derived from those data; and commitment to engaging in external assessment discussions (Rolheiser and Ross, 2001, cited in Campbell and Levin, 2008). As explained in Morris (2011), the literature stresses that for standardised test results to be used effectively, educators must have the capacity to assess, understand and apply such data. Without developing assessment capacity, the result can be "a sorry mixture of confusion, technical naivety and misleading advice" (Goldstein, 1999, cited in Campbell and Levin, 2008). In Ontario, Canada, developing capacity and assessment literacy is the responsibility of the school district. Campbell and Fullan (2006) found that school districts in Ontario that showed improved student outcomes also identified the development of assessment literacy at both the school and district levels as important activities (cited in Campbell and Levin, 2008). Such development activities included: providing professional development on data analysis and assessment literacy for principals and teachers; clearly setting expectations about the use of student assessment information; supporting schools in using and understanding data; encouraging the use of data to inform improvement planning, set goals and provide feedback (Campbell and Levin, 2008). Training benefits the assessment framework not only by providing teachers and school leaders with the specialised skills needed to utilise test results, but also by engaging them in the system thereby increasing stakeholder buy-in (Morris, 2011).

In Italy, the National Institute for the Evaluation of the Education System (INVALSI) is developing actions to promote the use of results from national standardised assessments among teachers and school principals in individual schools. INVALSI conceives a "School ID", summarising a school's performance in national student assessments together with contextual information with reference to provincial, regional, and state results, as a means to trigger dialogue among teachers, school principals and inspectors about what can be learned from student results (INVALSI, 2010).

Another area of priority is capacity for formative assessment. For instance, in Ireland, whole-school evaluation reports by inspectors and the 2009 National Assessments survey (Eivers et al., 2010) indicate that there is still significant scope for development in terms of the formative use of student assessment not only to improve the learning of individual students in the classroom but to promote improvement at whole-school level. These concerns have been instrumental in informing the new emphasis placed on evaluation and assessment in Ireland's 2011 National Literacy and Numeracy Strategy (Irish Department of Education and Skills, 2012).

Figure 3.10 Teachers' perceptions of professional development needs in student assessment practices (2007-08)

Percentage of teachers of lower secondary education indicating they have a "high level of need" for professional development in student assessment practices



StatLink http://dx.doi.org/10.1787/888932791267

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

A major concern in countries in which education governance is highly decentralised is the capacity of local education authorities to ensure the quality assurance of their schools. In many of these countries, it is recognised that many (often smaller) municipalities lack the capacity to develop and manage robust quality assurance systems, monitor schools effectively and follow up with schools accordingly. Further, the background and qualifications of officials in local education authorities responsible for school evaluation tend to vary significantly. This often leads to low expectations and the lack of an evaluation culture; variability and inconsistency in quality assurance practices; and limited capacity and skills of schools to use data.

In some countries schools are provided with the opportunity to hire specific expertise in evaluation and assessment through the availability of specifically trained and accredited experts in educational evaluation. Box 3.2 provides the examples of evaluation advisors in Denmark and evaluation experts in Hungary.

Box 3.2 Availability of evaluation experts in Denmark and Hungary

Evaluation advisors in Denmark

A development in the effort to strengthen the evaluation culture in Danish schools is the introduction of the role of evaluation advisor among the so-called "resource persons" that Danish schools can hire. Resource persons are teachers who undertake specific training and acquire expertise in a given domain who then perform the function of expert in that domain as part of their duties as a teacher (e.g. IT tutors, reading tutors, librarians). Evaluation advisors are still a limited resource in Danish schools: in a study from the Danish Evaluation Institute, only 8% of schools examined in the study had an evaluation advisor. The role of an evaluation advisor broadly consists of supporting the school effort in developing evaluation practices and an evaluation culture. It might involve the guidance and coaching of colleagues and school management on self-appraisal of teaching practices, peer feedback (including classroom observation), new knowledge and/or initiatives in the educational field, implementation of educational policies, co-ordination of quality assurance within the school or simple individual advice to teachers.

Source: Shewbridge et al. (2011b).

The National List of Experts in Hungary

In Hungary, the public education system organises an extensive professional service network, the National List of Experts, which serves as a pool of experts to be used by school maintainers in evaluation activities (e.g. school evaluation, strategies for school development). This list was first established by the Public Education Act of 1993, and it is the education administration's ongoing responsibility to co-ordinate the accreditation of experts to be part of the list. Experts on the list must participate in further education at least every five years. The Educational Authority, the agency which currently co-ordinates the list, is entitled to assess the activity of experts.

Source: Hungarian Ministry of Education and Culture (2010).

Skills for school leadership

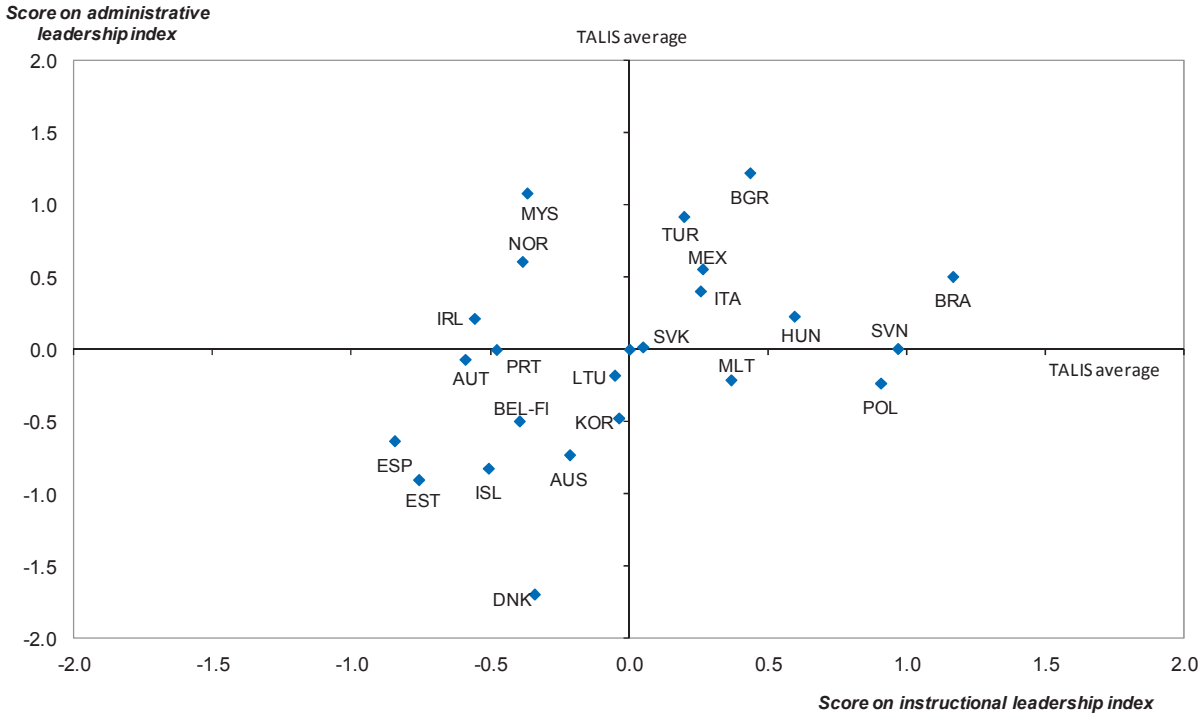
School leadership plays a key role not only in enhancing teaching and learning in schools but also in strengthening evaluation and assessment activities at different levels, including school self-evaluation processes and pedagogical guidance and coaching to individual teachers. As a result, an important policy lever is the development of evaluation and assessment competencies among school leaders. The work of school leaders is also crucial to building school capacity for accountability within the school community because accountability processes are nested in beliefs, experiences, and practices in schools (Hooge et al., 2012). It requires school leaders who are willing and able to empower staff, and in turn, to involve and share responsibility with parents and other interested members of the local community. It also requires school leaders who are willing to be held accountable by them (Leithwood, 2001).

In most countries, there is no specific initial education to train school leaders, nor does the specific career of school leader exist (see Chapter 7). Also, it is still common across the OECD area that school principals focus their work on administrative tasks, limiting their pedagogical leadership of schools. Figure 3.11 shows countries' relative positions in terms of the perceptions of school principals of their management styles, using specific indices based on TALIS data. Countries where school principals perceive administrative tasks are predominant in their work include Ireland, Italy, Mexico, Norway and Turkey. By contrast, in countries such as Australia, Denmark, Hungary, Iceland, Korea Poland and Slovenia instructional leadership seems to prevail over administrative tasks in the work of school principals, according to their perceptions. A large-scale

longitudinal study by Seashore Louis and colleagues (2010) based in the United States found that collective leadership focused on instructional improvement had a significant impact on teachers' working relationships, and on student achievement.

Figure 3.11 School principals according to their management styles (2007-08)

Scores on TALIS administrative leadership and instructional leadership indices



StatLink http://dx.doi.org/10.1787/888932791286

Definitions: *Instructional leadership* index: average of the indices for management-school goals, instructional management and direct supervision in the school; *Administrative leadership* index: average of the indices for accountable management and bureaucratic management.

Source: Reproduced from OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Tools and guidelines for evaluation and assessment

A typical strategy in countries to develop capacity for evaluation and assessment consists of efforts to build up a knowledge base, tools and guidelines to assist evaluation and assessment activities. The objective is to provide school agents with a comprehensive toolkit to engage in evaluation and assessment. Examples include:

- detailed guidelines to implement the curriculum
- marking rubrics listing criteria for assessing and rating different aspects of student performance and exemplars illustrating student performance at different levels of achievement
- range of optional assessment tools for teachers to use in student assessment, including formative assessment

- instruments to interpret results in student standardised assessments
- tools for school self-evaluation
- guidance for the application of teacher appraisal, including instruments for self-appraisal.

A number of systems such as New Zealand, Scotland and Sweden have developed "on-demand" assessments. Teachers may decide when students are ready to take a test in a particular subject or skill area, drawing from a central bank of assessment tasks. Control over the timing of tests means that teachers are able to provide students with feedback when it is relevant to the learning unit. In Scotland, a central system maps assessment tasks to standards and critical skills, topics and concepts in the curriculum. The assessments are usually designed, administered and scored locally, based on central guidelines and criteria.

In some countries, schools are supplied with statistical indicators intended to assist them in their self-evaluation. In Scotland, the Standard Tables and Charts (STACs) provide comparative information on attendance at school, as well as school costs per student and school leaver destinations. In Austria, the *Qualität in Schulen* and *Qualitätsinitiative Berufsbildung* Internet platforms supply schools with information and tools for both evaluation and data analysis, strengthening schools' capacity to self-evaluate. In Italy, schools receive regular feedback from the National Institute for the Evaluation of the Education System (INVALSI) on the performance of its students in national standardised assessments, including performance levels at the student, classroom and school level (across the different learning areas within a subject), variance within and across classes and information about specific disadvantaged groups.

Use of results

Knowledge management

Information systems

The overall evaluation and assessment framework produces large amounts of information and data which can subsequently be used for public information, policy planning and the improvement of practices across the education system. As analysed by Fazekas and Burns (2012), knowledge is crucial for governance and governance is indispensible for knowledge creation and dissemination. As complexity in education systems continues to increase, governance systems' capacity to learn becomes more and more crucial. Most institutions involved in education policy have become knowledge-intensive organisations whose success depends most critically on their ability to learn (Fazekas and Burns, 2012).

Making the best use of the evidence generated by evaluation and assessment activities depends to a large extent on the development of coherent information management systems. These include elements such as:

- standard framework for data collection and reporting
- data information/management systems – collection of data on students, teachers, schools and their performance over time

- public information – arrangements to share information about evaluation and assessment results with the general public
- identification of best practices and their dissemination across the system.

In Australia, there are standard frameworks both for reporting key performance measures (the *Measurement Framework for Schooling in Australia*, ACARA, 2010) and for general government sector reporting (the *Report on Government Services' Performance Indicator Framework*). This is in addition to the standardised Australian Bureau of Statistics National Schools Statistics Collection (NSSC) and the nationally comparable data on student outcomes (through the National Assessment Program). These entail the establishment of protocols to harmonise, standardise, and share the data among key stakeholders (DEEWR, 2010).

Some education jurisdictions in Australia have also developed sophisticated data information systems. One good example of practice is the School Measurement, Assessment and Reporting Toolkit (SMART) developed by New South Wales (see Box 3.3). This initiative has the potential to assist teachers in the instruction of their students, provide quick feedback to school agents, serve as a platform to post relevant instructional material to support teachers and improve knowledge management, operate as a network to connect teachers and schools with similar concerns, and create a better data infrastructure for educational research. In addition, schools' data management systems to track progress of individual students are also common in Australian schools. This means that the development of individual students can be tracked over time and that such information can be shared among teachers or with a student's next school (Santiago et al., 2011). Data collection and management systems sold by information technology companies are quite common in countries.

In Korea, data collection and management for evaluation and assessment is provided by three distinct systems which are linked together: the National Education Information System (NEIS), the School Information Disclosure System, and the statistical survey of education. NEIS comprises three types of data: academic affairs administration (as a support for teachers), general school administration (e.g. management of teaching body, school budget), and information for parents (e.g. student grades, attendance, school curriculum). The School Information Disclosure System provides a range of data and information to promote research, induce parental input into school education, and raise effectiveness and transparency in educational administration. It includes the results of the National Assessment of Educational Achievement and the results of the Teacher Appraisal for Professional Development (average grades per school). The statistical survey of education reviews and analyses basic education and administrative data and provides indicators at the metropolitan/provincial office level on characteristics of schools, teachers and school facilities (Kim et al., 2010).

In Norway, the development of the School Portal (*Skoleporten*) has been instrumental in ensuring access for school owners and schools to monitoring information and analyses of their results. The School Portal is a web-based information tool presenting key education monitoring information including learning outcomes, learning environments, resources and basic school data. The Portal has an open part accessible to the general public and a password-protected part where schools and school owners can access more detailed information and benchmark themselves against the national average. This approach holds promise for encouraging a more systematic and well-integrated way of using analyses of data in the process of self-evaluation and improvement planning. At the same time, it provides the general public with information on educational outcomes

(Norwegian Directorate for Education and Training, 2011). Similarly, in the French Community of Belgium, the Learning Portal (*Le Portail de l'Enseignement*) provides extensive information to the general public on student learning objectives, education indicators and education legislation as well as pedagogical tools and the instruments necessary for schools to engage in external evaluations (Blondin and Giot, 2011).

Box 3.3 Data information systems in Australia (New South Wales)

New South Wales SMART system

The New South Wales (NSW) Department of Education and Training has developed a sophisticated tool for data analysis in the form of the School Measurement, Assessment and Reporting Toolkit (SMART). This provides diagnostic information on NAPLAN (National Assessment Program – Literacy and Numeracy), ESSA (a Year 8 NSW-based science test) and the Higher School Certificate examination. This information, together with information from school-based assessment activities provides a wealth of objective diagnostic information to which teachers can respond. The SMART system is an example of how digital technology can assist in effectively using data and is now also used in the Australian Capital Territory and South Australia.

Analysis of educational outcomes and processes in NSW can be undertaken at many levels, from individual students, to groups of students, cohorts, schools and the system as a whole. The SMART package allows educators to identify areas for improvement as well as strengths in student performance. SMART also provides support through specific teaching strategies designed to improve student outcomes. SMART includes a number of functionalities intended to analyse NAPLAN results in-depth (see Table 19 in DEEWR, 2010).

For more information: *www.schools.nsw.edu.au/learning/7-12assessments/smart/index.php.*

Source: DEEWR (2010).

Identification of best practices

Another important aspect of knowledge management is to set up systematic processes to identify best practices within the overall evaluation and assessment framework and ensure that they are spread and shared across educational jurisdictions and schools. There is often a wide range of quality assurance activities developed locally within classrooms and schools, which tends not to be documented. A consequence is that the existing knowledge and information on evaluation and assessment may get lost and there is little systemic learning over time. In a decentralised system such as Norway's, networking is a common form of organisation among municipalities and there are a range of good examples where networks and partnerships have been established between different actors as a means to take collective responsibility for quality evaluation and improvement (Norwegian Directorate for Education and Training, 2011). Networks can be a powerful organisational tool embedding reform in the interactions of different stakeholders, sharing and dispersing responsibility and building capacity through the production of new knowledge and mutual learning that can feed back into policy and practice (Katz et al., 2009; Chapman and Aspin, 2003). Box 3.4 provides a number of examples of collaboration initiatives and partnerships in Norway.

Box 3.4 Collaboration initiatives and partnerships in Norway

In Norway, there are many examples of localised collaboration initiatives launched and developed by small clusters of municipalities as well as larger regional or national partnerships that are supported by the Association of Local and Regional Authorities (KS) or the Directorate for Education and Training. Examples are:

Municipal networks for efficiency and improvement: In 2002, the Association of Local and Regional Authorities (KS), the Ministry of Labour and Government Administration, and the Ministry of Local Government and Regional Development set up "municipal networks for efficiency and improvement" that offer quality monitoring tools for municipal use and provide a platform for municipalities to share experience, compare data and evaluate different ways of service delivery in different sectors. For the education sector, an agreement was established between KS and the Directorate for Education and Training to allow the networks to use results from the user surveys that are part of NKVS (the national quality assessment system). The networks bring together municipal staff and school leaders to discuss school evaluation and assessment issues and engage in benchmarking exercises. Each network meets four or five times and then the opportunity is offered to another group of municipalities.

Regional groups working on external school evaluation: The national school improvement project Knowledge Promotion – From Word to Deed (2006-2010) was launched by the Directorate for Education and Training to strengthen the sector's ability to evaluate its own results and plan improvement in line with the objectives in the Knowledge Promotion reform. One of the outcomes of the project was the establishment of 11 regional groups to work on external school evaluation. These groups received training in the programme's methodology for external school evaluation and have begun to establish local systems for external school evaluation.

Guidance Corps for school improvement: The Directorate has also recently established a "Guidance Corps" of exemplary school leaders who make themselves available to intervene in municipalities that have been targeted as needing help with capacity development (amongst others the municipalities from the "K-40" project). The "K-40 project" is a voluntary support offered to municipalities by the Directorate.

Collaboration of teacher education institutions and schools: An important recent development is the organisation of teacher education into five regions. This regionalisation of teacher education is intended to enhance the co-operation of teacher education institutions among each other and to develop partnerships between teacher education colleges, universities and schools. Every teacher education institution is required to participate and set up partnerships with local schools. While the Directorate for Education and Training has set up the infrastructure for this co-operation, it is now up to the participating institutions to take it further.

Source: Nusche et al. (2011b).

Innovation in education

The interplay between innovation in education and evaluation and assessment is complex and one which is difficult to analyse empirically. Nonetheless a number of potential linkages have been identified such as:

- Assessment as part of the process of innovation

  Those implementing innovations in education need to assess their effectiveness and to make necessary adaptations. Evidence on the impact of new approaches is also essential for successful dissemination (Looney, 2009). As such, assessment is vital for the process of validating innovations.

- Innovative assessment as a prerequisite for the implementation of pedagogical innovations

Evaluation and assessment procedures need to align with changes in what students should achieve as curriculum innovations occur. Given the strong backwash effect of assessment on learning (see Chapter 4), innovations in pedagogy are unlikely to be successful unless they are accompanied by related innovations in assessment (see Chapter 4). Innovative programmes also face additional barriers if assessment systems in place do not capture the innovative aspects of the programmes, missing important learning goals emphasised in those innovative programmes (Looney, 2009). Several innovations in assessment are taking advantage of recent advances in ICT to better respond to pedagogical innovations (see Chapter 4).

- *High-stakes* assessment as an inhibitor of educational innovation

  High-stakes assessments have the potential to undermine educational innovation. *High stakes* – such as publication of student assessment results at the school level or financial rewards for schools and teachers on the basis of student results – are intended to provide incentives for teachers and schools to focus on aspects measured, and to provide information for school improvement. Yet, these high stakes also discourage risk taking necessary for innovation, and may often encourage teachers to "teach to the test" (Looney, 2009). As put by Sawyer (2008), "The standards movement and the resulting high-stakes testing are increasing standardisation, at the same time that learning sciences and technology are making it possible for individual students to have customised learning experiences. Customisation combined with diverse knowledge sources enable students to learn different things. Schools will still need to measure learning for accountability purposes, but we do not yet know how to reconcile accountability with customised learning."

  It is interesting to note that some settings, typically characterised as *low stakes*, such as vocational education schools and alternative schools (e.g. Montessori, Steiner schools) seem to provide opportunities for innovation. Sliwka (2008) argues that alternative schooling has been influential in recent years as the instructional strategies and assessment techniques they have developed have impacted on teaching and learning in many public school systems across the world. Many of the so-called "authentic forms" of assessment that are used in mainstream education today originated in alternative schools (Sliwka, 2008). Also, in vocational education schools, the development of approaches to assessment focused on motivating students, giving high-quality feedback, and including the active participation of learners in the assessment process have influenced assessment practices in general education.

- Innovative assessment as leading to pedagogical innovations

  Evaluation and assessment can be a lever to drive innovation in education by signalling the types of learning that are valued (see Chapter 4). This is likely to be facilitated by some discretion at the local and school level to develop curriculum innovations and approaches to evaluation and assessment.

Evidence-based policy

The principle of informing policies and educational practices with evidence from research, including that generated by using the results of evaluation and assessment activities, is among the main goals of the evaluation and assessment framework. The

objective is to ensure that evaluation and assessment results are used in academic and policy research which subsequently informs the development of education policy. This includes developing evaluation and assessment policies which are evidence-based. Evidence-based policy includes aspects such as:

- the systematic use of evidence for policy development, including the use of evaluation and assessment results
- involvement of the research community in the use of results generated by the evaluation and assessment framework
- research units within national evaluation agencies, including the promotion of independent research and analysis
- strategic approach to research, analysis and evaluation by education authorities.

The principle of evidence-based policy making is well established in New Zealand. At the national level there is a strong commitment to bringing together national and international evidence on the factors and practices that can contribute to improving teaching and learning. The most prominent example is the Ministry of Education's Iterative Best Evidence Synthesis (BES) programme, which analyses research on school factors that have a positive effect on student learning. The publications appear to be widely used by both policy makers and stakeholder groups to inform education policy and practice in New Zealand. New Zealand researchers and academics also contribute regularly to debates on educational evaluation and assessment policies, both individually and collectively via advisory groups, the New Zealand Council for Educational Research (NZCER) and the recently created New Zealand Assessment Academy (NZAA). The New Zealand Teachers Council (NZTC) also contributes to building a sound evidence base on high-quality teaching. The Education Review Office (ERO) reviews international and national evidence on effective practice to underpin its methodology and indicators framework. ERO's evaluation indicators are informed by educational research, in particular the Best Evidence Syntheses described above and ERO's own evaluations of effective schools. In its publication on *Evaluation Indicators for School Reviews*, ERO provides a list of research studies that have informed each set of indicators (New Zealand Ministry of Education, 2010).

In the Netherlands, the Knowledge Directorate within the Ministry of Education was established to function as a clearing house of scientific knowledge and to stimulate the use of evidence in education policy development. Its role has benefitted from the renewed interest in and expansion of monitoring, assessment and evaluation procedures within the education system (Scheerens et al., 2012). In Slovenia, the Council for Quality and Evaluation is an advisory body composed of experts (typically researchers and academics) which co-ordinates evaluation processes in pre-tertiary education. Standards and procedures for evaluation and quality assurance are adopted by the Minister on the basis of the recommendations by the Council. The Council's tasks include the monitoring of the implementation of new educational programmes, the preparation of reports to inform decision making (by experts' councils and the Ministry of Education), the presentation of results to the general public, and proposals for future research work (Brejc et al., 2011).

In Denmark, the Council for Evaluation and Quality Development of Primary and Lower Secondary Education, as part of its mandate to monitor the academic level in the *Folkeskole* (compulsory education), commissions research and evaluation studies. The latter often include studies on the implementation and use of new national evaluation and assessment tools, e.g. Individual Student Plans and municipal quality reports and have led

to considerations and pilots of how to make such tools most relevant to local needs. One of the ten major challenges identified by the Council for the "*Folkeskole* 2020" is to strengthen the systematic exchange of knowledge between research institutions and schools to promote school use of knowledge to improve teaching. The Council aims to collect and disseminate research results to support the formation of policies for school improvement (Danish Ministry of Education and Rambøll, 2011). In the French Community of Belgium, the "Monitoring Commission" (*Commission de Pilotage*), among other missions, reviews research in education and ensures the relevant results inform policy development and school practices. In addition, it identifies research gaps in education and proposes to the government a multi-year research plan establishing research priorities in education as well as objectives to be reached by research studies (Blondin and Giot, 2011).

In Luxembourg, the Ministry of National Education and Vocational Training has involved research units from the University of Luxembourg in most of its reform efforts, particularly in the area of educational quality. These are namely, the EMACS (Educational Measurement and Applied Cognitive Science), INSIDE (Integrative Research Unit on Social and Individual Development), LCMI (Language, Culture, Media, Identities) and IPSE (Identities Politics, Societies, Spaces) (ADQS, 2011). In Austria, the concern of strengthening evidence-based policy contributed to the creation of the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE). BIFIE was given responsibility for the implementation and analysis of educational standards, the elaboration of a centralised competencies-based school exit examination (*Matura*), the establishment of an ongoing system of educational monitoring and the preparation of national education reports (Specht and Sobanski, 2012).

In Korea, a range of research institutes support education authorities (at the central and local levels) and schools with policy research and analyses of policy implementation. These include the Korean Educational Development Institute (KEDI), the Korean Institute for Curriculum and Evaluation (KICE), the Korean Research Institute of Vocational Education and Training (KRIVET), the Korean Education and Research Information Service (KERIS) and the National Institute for Lifelong Education (NILE) (Kim et al., 2010). Similarly, in France, a number of research institutes support the monitoring of the education system and the evaluation of education policies. These include the French Institute on Education (*Institut français de l'éducation*, IFE), the Research Institute on Education: Sociology and Economics of Education (*Institut de recherche sur l'éducation: Sociologie et économie de l'éducation*, IREDU) and the Centre for Studies and Research on Qualifications (*Centre d'études et de recherche sur les qualifications*, Céreq) (Dos Santos and Rakocevic, 2012).

Implementation of evaluation and assessment policies

A key challenge for policy makers is to move from knowing what changes are needed to implementing those changes successfully. Implementing educational evaluation policies is complex, involves a wide range of stakeholders with distinct interests, and requires informed debates and capacity building. Policies dictated at the national or local level are not always implemented at the school and classroom levels to the desired extent. Implementation difficulties may arise as a result of a wide range of factors, including lack of consensus among stakeholders, insufficient information and guidance to implement policies, limited involvement of professionals, inadequate competencies, narrow resources, scarce evidence basis or poor leadership to implement reforms.

Divergence of views and interests

To begin with, there might be significant divergences of views and interests among the relevant stakeholders as a result of the distinct perceived benefits and costs of policy initiatives. This diversity of views makes the policy making exercise particularly challenging, especially so given that policy makers often represent one of the stakeholder groups – the government authorities – and therefore they need to reconcile different perspectives to avoid the perception that evaluation and assessment policy is imposed to other groups in a top-down fashion. For example, in the choice of teacher appraisal methods, the relative importance of the summative and formative purposes is particularly contentious. On the one hand, policy makers and parents tend to value quality assurance and accountability. "They make the point that public schools are, after all, public institutions, supported by tax payer money, and that the public has a legitimate interest in the quality of the teaching that occurs there. It is through the system of teacher evaluation that members of the public, their legislators, local boards of education, and administrators, ensure the quality of teaching" (Danielson and McGreal, 2000). On the other hand, teachers and their unions expect opportunities of social recognition of their work and opportunities for professional growth through the development of a formative system of teacher appraisal (Avalos and Assael, 2006).

In the Czech Republic, a prominent example of implementation difficulties has been the introduction of the common standardised part of the school-leaving examination. Its development started in 1997 while its implementation occurred in spring 2011. In this long period several models were developed, pilot versions implemented, fundamental features modified several times (e.g. whether it should have one or two levels), and heated debates organised. The approach to the examination did not receive consensus among political parties and became an issue for political fights among some groups. This particular reform was characterised by significant uncertainty, a fragmentation of adaptations, and the dominance of politics to the detriment of pedagogical aspects (Santiago et al., 2012c).

Building consensus through consultation

A number of authors stress the importance of mechanisms to build consensus for successful adoption and implementation of policy initiatives (Fiske, 1996; Finlay et al., 1998; Corrales, 1999). There is extensive evidence that consensus is almost a prerequisite for successful implementation of policy reforms. As noted by Fiske (1996) with respect to school decentralisation, researchers are almost unanimous in arguing that if school decentralisation is going to be successfully carried out and have a positive impact on the quality of teaching and learning, it must be built on a foundation of broad consensus among the various actors involved and the various interest groups affected by such a change. And in fact, he observes on a basis of a comparative analysis that countries where leaders sought to build consensus for reform happen to be those where decentralisation was most successful. Building consensus is characterised by iterative processes of proposals and feedback which allow legitimate concerns to be taken into account, and hence reduce the likelihood of strong opposition by some stakeholder groups. A merit of structured consultations with stakeholders is that their regular involvement in policy design helps them build capacity over time. Another advantage is the potential for collective learning which might contribute to the development of a common concept of evaluation among stakeholders.

There is broad agreement in the literature that the involvement of stakeholders in education policy development cultivates a sense of joint ownership over policies, and hence helps build consensus over both the need and the relevance of reforms (Finlay et al., 1998; OECD, 2007). Policies promoting consensus build trust between the various stakeholder groups and policy makers. Keating (2011) analysed how various school districts in the United States developed and implemented new school principal appraisal systems. In most school districts, collaboration between different stakeholders (e.g. unions, teachers, school leaders and community representatives) played a key role in the design and implementation stages. The setting of shared priorities, negotiation, consensus building and transparency often resulted in greater ownership and acceptance among stakeholders.

The experience of countries participating in the Review suggests that mechanisms of regular and institutionalised consultation – which are inherent to consensual policy making – contribute to the development of trust among parties, and help them reach consensus. In Denmark, following the 2004 OECD recommendations on the need to establish an evaluation culture, all major stakeholder groups formed broad agreement on the importance of working to this end. Stakeholders worked together in a number of groups set up by the Minister of Education to come to agreement on how to follow up on the OECD recommendations and these were documented in "The *Folkeskole*'s response to the OECD" (Danish Ministry of Education and Rambøll, 2011). Box 3.5 outlines a range of initiatives in Denmark for promoting dialogue and reaching common views on educational evaluation policies.

Also, at the heart of the New Zealand education system is a strong trust in the professionalism of all actors and a culture of consultation and dialogue. Overall, the development of the national evaluation and assessment agenda has been characterised by strong collaborative work, as opposed to prescriptions being imposed from above. As a result of this participative approach, there appears to be considerable agreement and commitment of schools into overall evaluation and assessment strategies. While there are differences in views, there seems to be an underlying consensus on the purposes of evaluation and an expectation among stakeholders to participate in shaping the national agenda (Nusche et al., 2012). Similarly, policy making in Norway is characterised by a high level of respect for local ownership and this is evident in the development of the national evaluation and assessment framework. School owners and schools have a high degree of autonomy regarding school policies, curriculum development and evaluation and assessment. There is a shared understanding that democratic decision making and buy-in from those concerned by evaluation and assessment policy are essential for successful implementation (Nusche et al., 2011b).

In Finland, the objectives and priorities for educational evaluation are determined in the Education Evaluation Plan, which is devised by the Ministry of Education and Culture in collaboration with the Education Evaluation Council, the Higher Education Evaluation Council, the National Board of Education and other key groups. The members of the Education Evaluation Council represent the educational administration, education providers, teachers, students, employers, employees and researchers and thus can influence the aims and priorities of educational evaluation (Finnish Ministry of Education and Culture, forthcoming). In the French Community of Belgium a "Monitoring Commission" (*Commission de Pilotage*) has been given a key role in the monitoring of the education system. It has two main missions: it co-ordinates and reviews the coherence of the education system, and it follows the implementation of pedagogical reforms. Its membership reflects all the relevant actors in the education system: the school inspection, the school organisers, researchers, teacher unions and parent representatives (Blondin and Giot, 2011).

In Ireland, the involvement of stakeholders in the formulation of assessment and evaluation policy has been a prominent feature. A range of well-established frameworks promote dialogue and common action among the main stakeholders. The various committees that advise the assessment proposals of the National Council for Curriculum and Assessment (NCCA) and the national assessments of the Educational Research Centre (ERC) comprise members of relevant organisations and bodies including the Department of Education and Skills, school management groups, teacher unions, and parents' groups (Irish Department of Education and Skills, 2012). Nevertheless, it should be noted that criticisms have been made of the way in which consultation and consensus building have slowed down the development of policy and the implementation of radical change (Irish Department of Education and Skills, 2012).

Box 3.5 Promoting dialogue and reaching common views on educational evaluation in Denmark

In Denmark, there is a general tradition of involving the relevant interest groups in the development of policies for primary and lower secondary schools (*Folkeskole*). The key interest groups are diverse: Education authorities at the central level, municipalities (Local Government Denmark), teachers (Danish Union of Teachers), school leaders/principals (The Danish School Principals Union), parents (The National Parents' Association), students (Danish Students), the association for municipal management in the area of schools, associations representing the interests of the independent (private) primary schools in Denmark, and researchers. The Council for Evaluation and Quality Development of Primary and Lower Secondary Education is the most prominent platform for dialogue in relation to evaluation and assessment policies among these interest groups. It works on collecting and disseminating the most important research results to provide input to the policy process on school development. A range of other initiatives serve as platforms for promoting dialogue and reaching common views on education evaluation in Denmark. Examples include:

- A reference group was set up to guide the project "Strengthening of the evaluation culture in the *Folkeskole*". The reference group, whose membership includes all the relevant stakeholder groups, meets on a regular basis to discuss the project. This includes, for instance, the development of the national student tests.
- The interest groups of the *Folkeskole* were involved in 2010-11 in a committee established by the Minister of education aiming at deregulating the *Folkeskole*.
- In 2007-08, the Danish Union of Teachers and the Ministry of Education collaborated on a project called "The School of the Month". Each month, a school was celebrated for remarkable results. The project has since been pursued under the heading "the good example of the month" (*www.skolestyrelsen.dk*).
- The Local Government Denmark project "Partnership on the *Folkeskole*", involving 34 municipalities, has been a platform for co-operation and reflection between municipalities. (*www.kl.dk/ImageVault/Images/id_40353/ImageVaultHandler.aspx*).
- The Quality and Supervision Agency in collaboration with the Danish Evaluation Institute carry out "inspirational seminars" for teachers and school pedagogical staff with a view of encouraging schools to develop evaluation activities.
- The Quality and Supervision Agency has all major stakeholder groups represented in focus groups, which are being summoned on a regular basis to provide input on different initiatives related to the strengthening of the evaluation culture in the *Folkeskole*.
- The different interest groups of the *Folkeskole* launch on a regular basis common actions and/or common proposals related to issues in the *Folkeskole*, e.g. a paper with the title Common knowledge – Common action.

Source: Danish Ministry of Education and Rambøll (2011).

With respect to the initiation of new policies, the combination of top-down and bottom-up initiatives is generally believed to foster consensus (Finlay et al., 1998). For instance, a study of evidence-informed policy making underlines how the involvement of practitioners – teachers, other educational staff and their unions – in the production of research evidence and in its interpretation and translation into policy gives them a strong sense of ownership and strengthens their confidence in the reform process (OECD, 2007). In the Flemish Community of Belgium, the Flemish Education Council (VLOR) is an independent advisory body that the Ministry of Education and Training is required to consult when a draft decree is prepared for the Parliament. It brings together representatives from school organisers, school leaders, teachers, researchers, students and parents. One of the Council's activities is the organisation of consultations and conferences to discuss the state of education, including the policy implications of analysis of the results of national student assessments (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

Involvement of professionals

Involving teachers and school leaders in the development of education reforms is likely to facilitate their implementation. For example, by engaging teachers in the design, management and analysis of student standardised test results, teachers are more committed in the testing process and are more likely to apply the test results to improve student outcomes (Mons, 2009). Another example concerns the lead role to be played by teachers in developing and taking responsibility for teaching standards. Teachers' ownership of the teaching standards recognises their professionalism, the importance of their skills and experience and the extent of their responsibilities (Hess and West, 2006). Education authorities have also a lot to benefit from experienced teachers in providing advice for the design of teacher appraisal systems. Based on their own experience and research, they can be in a good position to provide expertise on what good teaching practices are and to help identify relevant criteria and instruments to evaluate teachers (Ingvarson et al., 2007). As a result, factors that influence the success of the introduction of an evaluation system include professionals' acceptance of the system and perceptions whether the evaluation processes are useful, objective and fair; and the extent to which evaluators and those being evaluated share a common understanding of evaluation purposes, procedures and uses.

Various researchers have stressed the importance of both, including the voices of stakeholders and professionals in the evaluation design process, as well as of including stakeholders and professionals in the evaluation procedures, as a precondition for establishing trust and collaboration (Clifford and Ross, 2011; Leon et al., 2011). Studies by Thomas et al. (2000) and Davis and Hensley (1999) on school leader and evaluator perceptions of school leader appraisal in Alberta, Canada, and California, United States, respectively revealed substantial differences between both groups, which provides some evidence for the importance of including school leaders in particular in all stages of the development process of appraisal systems as well as the appraisal process itself.

The involvement of teachers and school leaders in their own appraisal process has been identified as another key aspect for the successful implementation of individual appraisal processes. Engaging teachers and school leaders in their own appraisal, e.g. through setting objectives, self-appraisal and the preparation of individual portfolios, can help create a more effective and empowering process for teachers and school leaders, and, therefore, aid successful implementation.

Kennedy (2005) argues that highly dedicated teachers' reform rejections do not come from their unwillingness to change or improve, but from "the sad fact that most reforms don't acknowledge the realities of classroom teaching". Also, imposed change is believed to create a "culture of compliance" (Datnow and Castellano, 2000; Leithwood et al., 2002). Teachers who are constrained in ways likely to reduce their own intrinsic motivation to teach may behave in more controlling ways and be less effective in teaching their students. By contrast, if teachers are involved in planning and implementing evaluation schemes, they are more likely to sustain reform efforts (Leithwood et al., 2000).

Clarity of purposes

Another factor which is often put forward by researchers when analysing the reasons for the success or otherwise of policy adoption and implementation relates to the communication of the objectives and purposes of reforms. Indeed, Olsen (1989) notes that policies are more likely to succeed if their intentions are focused and well defined rather than ambiguous. This highlights the importance of clearly communicating the rationale of evaluation and assessment activities, the objectives they seek to achieve and their usefulness and value for the different stakeholders.

For instance, in the case of student standardised assessment, in order to promote desired responses on the part of teachers, it is critical that they understand and support the assessment goals (Hamilton and Stecher, 2002; Alliance for Excellent Education, 2010). Establishing clear goals and standards and communicating them to teachers mitigates strategic behaviour such as "teaching to the test" as teachers have a clearer sense of what they should be aiming for with regard to student outcomes (Hamilton and Stecher, 2002).

In Norway, unclear communication about what the National Quality Assessment System (NKVS) does – its objectives and how data from different quality assessments are integrated as a coherent whole – has meant that NKVS is yet to be fully accepted as a useful tool at the local level (Allerup et al., 2009). It is important that schools and school owners see NKVS as a tool they can use (from a learning perspective) and not just as a useful tool for national authorities to use for the monitoring of learning outcomes.

Evidence to inform consensus-building

Resistance to reform might also be due to imperfect information of stakeholders – either on the nature of the proposed policy changes, their impact, and most importantly, information on whether or not they will be better or worse off at the individual or group level. This also involves insufficient preparedness of the public opinion for some reforms and the resulting lack of social acceptance for policy innovations. This might be exacerbated by an underdeveloped culture and little tradition of evaluation in education.

This highlights the importance of promoting research and making the evidence underlying the policy proposals available to the relevant stakeholders in helping convince practitioners and society at large. The objective is to raise awareness on problematic issues, enhance the national debate and disseminate evidence on the effectiveness and impact of different policy alternatives, and hence to find a consensus on educational evaluation policy. In the case of teacher appraisal, Milanowski and Heneman (2001) found that teachers' overall favourableness toward a system newly implemented in a medium-sized school district in the United States was correlated with acceptance of the teaching standards, the perceived fairness of the process, the qualities of the evaluator, and the perception that the evaluation system has a positive impact on their teaching.

The concern of the Portuguese authorities to build teacher appraisal on research evidence and recognised good practice was a clear strength of the system when it was introduced. In 2007, the then Ministry of Education set up the Scientific Council for Teacher Evaluation (CCAP) as a consultative body to supervise and monitor the implementation of teacher appraisal (in late 2011, following the rationalisation of education services, the CCAP ceased its functions). The CCAP brought together educational researchers and distinguished teachers and as such was in a good position to recognise good evaluation practices, be informed of relevant research developments and provide evidence-based advice (Santiago et al., 2009).

In Hungary, the Council for the Evaluation of Public Education, established in 2004, is an advisory body of the Minister of Education and Culture which seeks to bring scientific evidence to the decision-making process within education. Its members are invited by the Minister of Education and Culture from among the most prestigious national and international academic experts in areas such as the appraisal of teacher effectiveness, measurement theory, data collection and data analysis, content framework development and the management of evaluation programmes. The Council submits proposals for the development of evaluation and assessment in Hungary (Hungarian Ministry of Education and Culture, 2010).[3]

Policy experimentation and pilots

Policy experimentation and the recourse to pilot schemes can prove powerful in testing out policy initiatives and – by virtue of their temporary nature and limited scope – overcoming fears and resistances by specific groups of stakeholders. A pilot implementation is a cost effective way to ensure that a given initiative meets its intended purposes before full implementation. Seeking feedback from the involved school agents during the pilot implementation is essential to correct the potential flaws and concerns related to the initiative being tested.

In Ireland, pilot projects are usually developed before wide-scale implementation. This is reflected in a school self-evaluation pilot project undertaken in 2010/11 by a sample of 12 primary schools in conjunction with the Department of Education and Skills. Similarly, the Project Maths initiative for second level schools began in September 2008, with an initial group of 24 schools. Project Maths involves the introduction of revised syllabuses for both Junior and Leaving Certificate Mathematics. It involves changes to what students learn in mathematics, how they learn it and how they will be assessed. The pilot project helps the National Council for Curriculum and Assessment (NCCA) to learn from schools how the proposed revisions to the syllabus work in classrooms and will lead to the development of teaching and learning resources and assessment instruments (Irish Department of Education and Skills, 2012).

Evaluation of implementation

Another approach is to periodically review and evaluate processes after the full implementation. Education professionals such as teachers and school leaders are more likely to accept a policy initiative today if they know that they will be able to express their concerns and provide advice on the necessary adjustments as the initiative evolves. Amsterdam et al. (2003) analysed the three-year development and validation of a school principal appraisal system (i.e. standards, criteria and instruments) in South Carolina, United States, that involved researchers from the South Carolina Educational Policy Center at the University of South Carolina, the South Carolina Department of Education,

a stakeholder committee (e.g. superintendents, school principals, teachers, guidance counsellors and journalists) and an expert panel. Superintendents responsible for carrying out the appraisal and school principals had the opportunity to further inform the development of standards and criteria through a survey and an online discussion group. The new standards, appraisal criteria and instruments subsequently underwent a process of piloting and validation through focus groups and a school principal survey of participants in the pilot to identify strengths and weaknesses of the new system. Based on their experiences, Amsterdam et al. argued that stakeholder input may help ensure that appraisal systems are practical and useful for those concerned, and that the appraisal is supported by key stakeholders. At the same time, the involvement of school leaders in the design of standards and appraisal criteria may help to establish an understanding of the aspects and criteria that school leaders will be appraised against.

In New Zealand, the Ministry of Education commissions independent evaluations to monitor the implementation of national policies. Examples are evaluations of the implementation of The New Zealand Curriculum and the National Standards. The implementation of The New Zealand Curriculum in English-medium schools was monitored by the Education Review Office (ERO) in a series of reports. The Ministry, in collaboration with ERO, also developed a framework to monitor and evaluate the implementation of National Standards. The National Standards: School Sample Monitoring and Evaluation Project, run by a contracted evaluation team, collects information from a sample of state schools over the period of 2009-13. This information is complemented by survey data, information from ERO reports and results from national and international assessments (Nusche et al., 2012).

In a range of countries, it is typical for external evaluation providers to collect feedback from schools and other stakeholders on their experience with the external evaluation process in order to monitor its implementation. School evaluation procedures may also be evaluated through national audits, stakeholder surveys, independent evaluations and research studies (see Chapter 6). The same happens in the area of teacher appraisal. For example, the state of Rhode Island in the United States has developed a formal mechanism for evaluating districts' teacher evaluation systems and using the resulting information for the continuous improvement and increasing validity of those systems. It builds on a sophisticated set of standards which are used to guide the evaluation of educator evaluation systems. The results of evaluations are used to continually refine instruments and processes over time as new information is collected and analysed. The six standards that comprise the Educator Evaluation System Standards support the work of school districts to assure educator quality through a comprehensive district educator evaluation system that: *(i)* establishes a common understanding of expectations for educator quality within the district; *(ii)* emphasises the professional growth and continuous improvement of individual educators; *(iii)* creates an organisational approach to the collective professional growth and continuous improvement of groups of educators to support district goals; *(iv)* provides quality assurance for the performance of all district educators; *(v)* assures fair, accurate, and consistent evaluations; and *(vi)* provides district educators a role in guiding the ongoing system development in response to systematic feedback and changing district needs (Rhode Island Department of Elementary and Secondary Education, 2009).

Capacity building

Fiske (1996) underlines the importance of training policies for effective and successful implementation, as a means to ensure that all stakeholders are equipped and

prepared to take on the new roles and responsibilities that are required from them as a result of education reforms. Limited professional expertise of those with responsibility to evaluate and little analytical capacity to use the results from evaluation and assessment have the potential to harm the implementation of evaluation and assessment policies. For instance, in teacher appraisal models, it is fundamental to provide in-depth training to evaluators to guarantee that they are legitimate in the eyes of teachers. Also, scepticism towards data among educators resulting from a lack of capacity of schools and teachers to understand and use data effectively to inform development is also likely to increase implementation difficulties (Campbell and Levin, 2008).

In Portugal, the implementation of teacher appraisal, introduced in 2007, has been challenging and has exposed a range of difficulties. These resulted from putting into operation a comprehensive model in a short time span and the little anticipation by government of the difficulties. There was little experience with and tradition of evaluation, the system was unprepared to undertake large-scale teacher appraisal as a result of the limited professional expertise of those with responsibility to evaluate, a sense of unfairness by those being evaluated emerged, excessive bureaucratic demands on schools were made, and little time was given to implement the model (for further details see Santiago et al., 2009).

A common challenge in decentralised countries, where local decision making is significant, is the limited capacity at the local level to implement evaluation and assessment policies. For example, there is considerable disparity in educational expertise across the school administration departments of the 430 municipalities in Norway. Smaller municipalities do not benefit from the same capacity to run quality assurance frameworks within their jurisdiction. There are indications that the requirement under the Knowledge Promotion Reform that the local authorities work on curricula and assessment is too demanding for municipalities (school owners), particularly for the smallest ones (Norwegian Directorate for Education and Training, 2011).

Resources

Educational evaluation initiatives require time and other resources for the school agents involved. In the case of teacher appraisal, Milanowski and Heneman (2001) found that even if teachers accept the standards and the need for an appraisal system, they may still manifest reluctance when the system adds too much to their workloads. As emphasised by Heneman et al. (2006), "System designers need to carefully review what is required of teachers to minimise burden. […] Perhaps some small reduction in other responsibilities while teachers are undergoing evaluation would decrease the perception of burden and sense of stress." Also in the context of teacher appraisal, Marshall (2005) indicates that policy makers should also aim at reducing the administrative workload for evaluators, especially school principals, in order to provide them with more time for teacher evaluation, feedback and coaching.

Research has highlighted the burgeoning workload many school leaders face in various countries (e.g. Pont et al., 2008). In England, for example, a study by PriceWaterhouseCoopers (2007) indicated that 61% of the school principals that took part in the study described their work-life balance as poor or very poor. A study by Cullen (1997) on the experience of school principal appraisal in England identified school principals' lack of time as one of the key challenges for ensuring successful implementation of school leader appraisal. In light of these insights, policy makers face the challenge of developing school leader appraisal processes that do not require an

excessive investment of time and efforts and that school leaders perceive and experience as meaningful and useful.

Timing

Timing is relevant to education reform implementation in a number of ways. To begin with, there is a substantial gap between the time at which the initial cost of reform is incurred, and the time when it is evident whether the intended benefits of reforms actually materialise. This makes reform a thankless task when elections take place before the benefits are realised. This, too, is a factor that complicates the politics of reform in many domains, but again, it seems to be of exceptional importance in education, where the lags involved are far longer than is typical of, for example, labour- or product-market reforms. As a result, the political cycle considerably conditions the timing, scope and content of education reforms. Timing can be important also with regard to the sequencing of different components of reform, if one element – curriculum reform, for example – requires prior reforms in pre-service and in-service training in order to be effective (Wurzburg, 2010).

Another important consideration is the need for policy reform to be tailored to the particular stage of development of the policy area being addressed. Time is needed to learn and understand, to build trust and develop the necessary capacity to move onto the next stage of policy development. For instance, work on student assessment by the World Bank distinguishes four phases of development of student assessment frameworks: latent (absence of assessment activity), emerging (enabling contexts, system alignment and assessment quality taking shape), established (enabling contexts, system alignment and assessment quality stable, assured, or consolidated in nature) and advanced (enabling contexts, system alignment and assessment quality highly developed in nature) (Clarke, 2012). This work highlights the importance of understanding how to progress through these phases of development and of designing policies which take into account the initial stage of development.

Pointers for future policy development

This chapter has reviewed country approaches to co-ordinating evaluation and assessment activities within their educational systems and developing evaluation and assessment frameworks. The policy suggestions that follow are drawn from the experiences reported in the Country Background Reports, the analyses of external review teams in Country Reviews, and the available research literature. It should be stressed that there is no single model or global best practice of bringing together an evaluation and assessment framework. The development of policies and practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for other countries they may have less relevance because of different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing policies for building and evaluation and assessment framework. The implications also need to be treated with caution because in some instances there is not a strong enough research base across a sufficient number of countries to be confident about successful implementation. Rather, the discussion attempts to distil potentially useful ideas and lessons from the experiences of countries that have been searching for better ways to frame educational evaluation and assessment.

Governance

Integrate the evaluation and assessment framework

Authentic evaluation, that which leads to the improvement of educational practices at all levels, is central to establishing a high-performing education system. It is also instrumental in recognising and rewarding the work of educational practitioners. Promoting evaluation and assessment is clearly in the national interest. As a result, more and more countries embark on ambitious school reform programmes which include a strong element of evaluation and assessment. This consolidates the evaluation culture in education systems and reinforces the role of evaluation and assessment frameworks in driving the reform agenda. However, the full potential of evaluation and assessment will not be realised until the framework is fully integrated and is perceived as a coherent whole. This requires a holistic approach to building a complete evaluation and assessment framework in view of generating synergies between its components, avoiding duplication of procedures and preventing inconsistency of objectives.

At the outset, it might prove useful to develop a strategy or framework document that conceptualises a complete evaluation and assessment framework and articulates ways to achieve the coherence between its different components. Voices of key stakeholders groups should be engaged in the development of the strategy so as to ensure that it is responsive to broader social and economic needs as well as to the goals of the education system. The strategy should essentially constitute a common framework of reference for educational evaluation across the country with the ultimate objective of embedding evaluation as an ongoing and essential part of the professionalism of the actors in the education system.

The strategy should establish a clear rationale for evaluation and assessment and a compelling narrative about how evaluation and assessment align with the different elements in the education reform programme. It should clearly communicate that the purpose of the evaluation and assessment framework is to improve the educational outcomes of students. As such, it is expected that school agents actively use the results of evaluation and assessment activities to develop improvement or action plans at all levels. The strategy should describe how each component of the evaluation and assessment framework can produce results that are useful for classroom practice and school development activities.

The strategy could also contribute to clarifying responsibilities of different actors for the different components and allow for better networking and connections between the people working on evaluation and assessment activities. As such, it should also create the conditions for a better articulation between the different levels of educational governance, including evaluation agencies and local education authorities. Finally, it is important that the strategy establishes linkages to evaluation in the overall public sector.

Align the evaluation and assessment framework with educational goals and student learning objectives

A critical aspect in the effectiveness of the evaluation and assessment framework is its proper alignment with educational goals and student learning objectives. This involves a range of aspects. First, it requires a given orientation for evaluation and assessment procedures to align with the main principles embedded in educational goals and student learning objectives. For instance, if educational goals are based on principles such as

student-centred learning, collaborative work, achievement of competencies and assessment for learning then there should be greater emphasis on the developmental function of evaluation and assessment, involving more attention to student formative assessment, greater emphasis on self-reflection for all the school agents, greater focus on continuous improvement in teacher appraisal, and better use of results for feedback.

Second, evaluation and assessment procedures require direct alignment with student learning objectives. This implies designing fit-for-purpose student assessments which focus on the competencies promoted in student learning objectives, ensuring the overall evaluation and assessment framework captures the whole range of student learning objectives, and developing teaching and school management standards which are aligned with student learning objectives. This could involve research and development to strengthen the range of measurement technologies available to assess, for instance, students' higher order skills such as problem solving, reasoning and communication (see Chapter 4).

Third, it is essential that all school agents have a clear understanding of education goals. This requires goals to be clearly articulated; the development of clear learning expectations and criteria to assess achievement of learning objectives; room for schools to exercise some autonomy in adapting learning objectives to their local needs; and collaboration among teachers and schools to ensure moderation processes which enhance the consistency with which learning goals are achieved. This should go alongside the kinds of supports and incentives for school agents to gain professional knowledge of the implications of educational goals for teaching, learning, evaluation and assessment. A prerequisite is to ensure that student learning objectives are grounded in evidence of how students learn and progress within and across different subject domains, and represent realistic goals for attainment.

Fourth, it is essential to evaluate the impact of evaluation and assessment against student learning objectives on the quality of the teaching and learning. Particular attention should be given to identifying unintended effects as evaluation and assessment activities have considerable potential to determine the behaviour of school agents. For instance, undesired effects such as teaching to the test and the narrowing of the curriculum have been identified as consequences of high-stakes assessments. Significant investments in research and development on the alignment of evaluation and assessment with student learning objectives should also be made as systems that are not well aligned waste significant resources.

Secure links to the classroom and draw on teacher professionalism

Realising the full potential of the overall evaluation and assessment framework involves establishing strategies to strengthen the linkages to classroom practice, where the improvement of student learning takes place. Evaluation and assessment have no value if they do not lead to the improvement of classroom practice and student learning. This calls for an articulation of ways for the evaluation and assessment framework to generate improvements in classroom practice through the assessment and evaluation procedures which are closer to the place of learning.

An important step in this direction could be a national reflection about the nature and purpose of evaluation components such as school evaluation, school leader appraisal, teacher appraisal and student formative assessment within the overall education reform strategy and the best approaches for these evaluation components to improve classroom practices. This reflection would shed light on strategies which can contribute to reinforce

the linkages between evaluation and assessment and classroom practice. Impacting classroom practice is likely to require the evaluation and assessment framework to place considerable emphasis on its developmental function. Channels which are likely to reinforce links to classroom practice include: an emphasis on teacher appraisal for the continuous improvement of teaching practices; ensuring teaching standards are aligned with student learning objectives; involving teachers in school evaluation, in particular through conceiving school self-evaluation as a collective process with responsibilities for teachers; ensuring that teachers are seen as the main experts not only in instructing but also in assessing their students, so teachers feel the ownership of student assessment and accept it as an integral part of teaching and learning; building teacher capacity for student formative assessment; and building teachers' ability to assess against educational standards.

The central agent in securing links between the evaluation and assessment framework and the classroom is the teacher. This highlights the importance for evaluation and assessment frameworks to draw on the professionalism of teachers in ensuring evaluation and assessment activities result in authentic improvement of classroom practices and student learning.

Give a prominent role to independent evaluation agencies

The governance of the evaluation and assessment framework could benefit from the existence of independent evaluation agencies. This would be in a context where education authorities retain the leadership in setting educational strategy and developing educational policy, and maintain a role in the implementation of all the components of the evaluation and assessment framework. Such independent evaluation agencies, which could take responsibility for areas such as the curriculum, assessment, reporting on student achievement, school performance and improvement, teaching and school leadership management, would take the lead in evaluation and assessment in the respective areas of responsibility, involve the range of relevant stakeholders in their activities, and provide an independent judgement of the achievement of education goals.

The establishment of these independent agencies would involve political and financial independence from education authorities and a significant presence of experts and specialists in their decision-making bodies. The objective would be to establish these agencies as authoritative voices in the areas they cover, highly credible for their expertise and technical capacity, and issuing recommendations for the implementation of evaluation and assessment procedures in the country. In terms of functions, these agencies should emphasise their technical leadership (e.g. in developing evaluation instruments, guidelines); the monitoring of the education system, the teaching and school leadership professions; the introduction of innovations on the basis of research results; the development of capacity for evaluation and assessment across the system; and their technical support for school agents to implement evaluation and assessment procedures at the local level. One such agency could become the entity with the responsibility to assess the state of education in the respective country and develop analysis to inform policy development by education authorities. It is also expected that the agencies' work is done in close dialogue with education authorities at all levels.

Promote national consistency while giving room for local diversity

In order to contribute to national reform agendas, a certain degree of national consistency of approaches to evaluation and assessment is desirable. This is likely to

provide greater guarantees that evaluation and assessment practices are aligned with national student learning objectives. However, in certain countries, there are strong traditions of local ownership – at the jurisdiction level (federal systems), local level (region or municipality), or school level. In these cases, a high degree of autonomy is granted in school policies, curriculum development and evaluation and assessment. There is an understanding that shared or autonomous decision making and buy-in from those concerned are essential for the successful implementation of evaluation and assessment policy. It is also clear that local actors are in a better position to adapt evaluation and assessment policies to local needs.

Hence, the evaluation and assessment framework will need to find the right balance between national consistency and local diversity. A possible approach is to agree general principles for the operation of procedures such as school evaluation, teacher appraisal, school leader appraisal and student assessment while allowing flexibility of approach within the agreed parameters to better meet local needs. For each of the evaluation components on which principles would be agreed, a number of fundamental issues could be addressed, such as: how to combine the accountability and developmental functions; the scope in relation to the national agenda; aspects to be assessed; reference standards; the role and nature of externality; and the extent of transparency. The principles agreed should come along with clear goals, a range of tools and guidelines for implementation. They should permit better consistency of evaluation practices across schools while leaving sufficient room for local adaptation.

In decentralised systems, it is also important to encourage the different actors to co-operate, share and spread good practice and thereby facilitate system learning, development and improvement. In some countries, networking and partnerships are common forms of organisation among schools or local education providers (such as municipalities) to take collective responsibility for quality evaluation and improvement.

Integrate the non-public sector in the overall evaluation and assessment framework

Evaluation and assessment practices in the non-public sector can be very diverse and display limited alignment with those in place in public schools. As a result, in spite of possibly well-consolidated practices in the non-public sector, there is limited guarantee that those practices are aligned with the national education agenda. There are a range of possible approaches to better integrate the non-public sector in the overall evaluation and assessment framework. One possibility is to require the non-public sector to comply with the approaches followed within the evaluation and assessment framework, especially for those sectors or schools which receive public subsidies. Another possibility is for the non-public sector to be part of protocol agreements which specify general principles for the operation of procedures such as school evaluation, teacher appraisal or the appraisal of school leaders while allowing flexibility of approach within the agreed parameters. The degree of integration of the non-public school sector within the evaluation and assessment framework should relate to the extent to which it receives public subsidies; recognise the degree of market-based accountability non-public schools are exposed to; and respect its freedom of organisation. At the system level, and in order to monitor their performance, non-public schools could be compelled to adhere to public administrative data collections and be part of common performance reporting for schools in all sectors. This would facilitate the reporting of comparable information across schools, which can greatly assist parental choice of schools. The adherence of non-public schools to common performance

reporting is particularly pertinent for schools which receive public subsidies as a way to hold them accountable for the use of public funds.

Design and procedures

Ensure core components are sufficiently developed within the evaluation and assessment framework

A priority is to ensure that the key components of the evaluation and assessment framework are sufficiently developed and contribute effectively to the overall evaluation and assessment strategy. A range of areas need reinforcement in some countries. For example, greater emphasis is frequently needed in consolidating student formative assessment and criterion-based student summative assessment by teachers. The latter often requires better moderation processes to ensure the consistency of student marking by teachers, a key area to guarantee fairness of student marking across schools in a given country.

In a range of countries, teacher appraisal also requires considerable policy attention. Processes are often not systematic enough to ensure that all teachers are appraised and subsequently receive feedback, professional development opportunities, and prospects of career advancement. This could involve a need to re-conceptualise teacher appraisal, develop teaching standards and provide a structure to support its implementation at the school level. Also, in some countries, greater incentives need to be provided to schools to engage in self-evaluation so it is systematically performed with the involvement of all schools agents and follow-up which leads to school improvement. This is to be complemented with requirements for external school evaluation, an exercise to be led by dedicated structures that have the capacity to support school development. Another area which could benefit from greater policy attention, underdeveloped in many countries, is the appraisal of school leaders. Finally, another typical area for further investment in countries is qualitative evaluation at the system level. Specific policy suggestions to develop these evaluation and assessment components are proposed in subsequent chapters.

Establish articulations between components of the evaluation and assessment framework

The process of developing an effective evaluation and assessment framework should give due attention to: achieving proper articulation between the different evaluation components (e.g. school evaluation and teacher appraisal); warranting the several elements within an evaluation component are sufficiently linked (e.g. teaching standards and teacher appraisal); and ensuring processes are in place to guarantee the consistent application of evaluation and assessment procedures (e.g. consistency of teachers' marks). Examples of articulations which are desirable to establish between components of the evaluation and assessment framework were given earlier in this chapter.

A prominent example is the articulation between school evaluation and teacher appraisal. Given that the systems of school evaluation and teacher appraisal and feedback have both the objective of maintaining standards and improving student performance, there are likely to be great benefits from the synergies between school evaluation and teacher appraisal. To achieve the greatest impact, the focus of school evaluation should either be linked to or have an effect on the focus of teacher appraisal. This indicates that school evaluation should comprise the monitoring of the quality of teaching and learning,

possibly include the external validation of school-based processes for teacher appraisal (holding school leaders accountable as necessary), and school development processes should explore links to the evaluation of teaching practice. In the context of school self-evaluation, it is also important to ensure the centrality of the evaluation of teaching quality and the feedback to individual teachers.

Place the students at the centre of the evaluation and assessment framework

Given that the fundamental purpose of evaluation and assessment is to improve the learning of the students, a key principle is to place the students at the centre of the framework. This translates into teaching, learning and assessment approaches which focus on students' authentic learning. Students should be fully engaged with their learning, contributing to the planning and organisation of lessons, having learning expectations communicated to them, assessing their learning and that of their peers, and benefitting from individualised support and differentiated learning. In addition, it is important to build community and parental involvement and an acceptance of learning and teaching as a shared responsibility. A particularly important priority for some countries is to reduce the high rates of grade repetition. There are alternative ways of supporting those with learning difficulties in the classroom. One way is to provide extra teaching time for students who fall behind and adapt teaching to their needs. There can also be short-term, intensive interventions of one-on-one lessons for underperforming students. This can be organised with extra staff such as recovery teachers (see also Field et al., 2007 and OECD, 2012b).

In addition, evaluation and assessment should focus on improving student outcomes and achieving student learning objectives. This should be reflected in the priorities for national monitoring, the importance of evidence on student performance for school evaluation and teacher appraisal, the value of clear reporting on student results, and the emphasis on feedback for improving student learning strategies. There is also the increasing recognition that the monitoring of student outcomes must extend beyond knowledge skills in key subject areas and include broader learning outcomes, including students' critical thinking skills, social competencies, engagement with learning and overall well-being (see also Chapter 4).

Build on some key principles to effectively implement evaluation and assessment

The strategy to develop an effective evaluation and assessment framework should build on some key principles, including:

- **The centrality of teaching and learning**: It is critical to ensure that the evaluation of teaching and learning quality is central to the evaluation framework. Classroom observation should be a key element of teacher appraisal as well as an important instrument in external school evaluation. Similarly, the observation of teaching and feedback to individual teachers should be part of school self-evaluation processes. The effectiveness of the evaluation and assessment framework will depend to a great extent on the ability to cultivate a culture of sharing classroom practice, professional feedback and peer learning.
- **The importance of school leadership**: The effective operation of evaluation and assessment will depend to a great extent on the way the concept and practice of school leadership gains ground within the education system. It is difficult to envisage either effective teacher appraisal or productive school self review without strong leadership capacity. It is essential that school principals take direct

responsibility for exerting instructional leadership and for assuming the quality of education in their schools. Hence, the recruitment, appraisal, development and support for school leaders is of key importance in creating and sustaining effective evaluation and assessment cultures within schools.

- **Equity as a key dimension in the evaluation and assessment framework**: It is essential that evaluation and assessment contribute to advancing the equity goals of education systems. At the system level, it is imperative to identify educational disadvantage and understand its impact on student performance. Developing equity measures should be a priority in all countries. It is also important to ensure that evaluation and assessment procedures are fair to given groups such as cultural minorities and students with special needs.
- **A structure to integrate accountability and development**: The overall evaluation and assessment framework should include elements to accomplish both the accountability and developmental functions at all levels of the system (e.g. formative vs. summative assessment for students; professional development for teachers vs. career advancement decisions following teacher appraisal; data reporting vs. improvement action plans for schools) and provide a structure which can potentially integrate these two functions.
- **Commitment to transparency**: The overall evaluation and assessment framework can be strengthened by a high level of transparency in monitoring and publishing results.

Capacity

Sustain efforts to improve capacity for evaluation and assessment

The development of an effective evaluation and assessment framework involves considerable investment in developing competencies and skills for evaluation and assessment at all levels. Hence, an area of policy priority is sustaining efforts to improve the capacity for evaluation and assessment. Depending on country specific circumstances, areas of priority might be: developing teachers' capacity to assess against student learning objectives; improving the skills of teachers for formative assessment; improving the data handling skills of school agents; and facilitating the understanding by parents and other stakeholders of the concepts behind the ways the data are presented and compared. Another area which deserves attention relates to skills and competencies for teacher appraisal and school evaluation. Capacity building through adequate provision of initial teacher education and professional development should be a priority making sure provision is well aligned with the national education agenda. This should go alongside the development of training and competency descriptions for key people within the evaluation and assessment framework.

There is also a need to reinforce the instructional leadership skills of school principals as their role in many countries still retains a more traditional focus on administrative tasks. The objective is that school leaders operate effective feedback, coaching and appraisal arrangements for their staff and effectively lead whole-school evaluation processes. This can primarily be achieved by redefining school leadership as educational leadership, and ensuring that the whole cohort of school leaders receives adequate training in "leadership for learning". School leaders should be trained to implement an

authentic evaluation of teaching and learning, feedback and objective setting at their schools, including techniques in teacher observation (see also Chapters 5 and 7).

There also needs to be strong capability at the national level to steer evaluation and assessment. This can be ensured through the establishment of agencies with high levels of expertise which have the capacity to foster the development of skills for evaluation and assessment across the system. Such agencies, as suggested earlier, could provide important leadership in modelling and disseminating good practice within the evaluation and assessment framework (see also Chapter 8).

Improve the articulation between levels of authority and assure support from the centre

There is a need to ensure a good articulation between the different levels of authority in the implementation of policies for evaluation and assessment (national, regional, municipal, non-public, school level). In addition to the regulatory provisions defining the respective responsibilities in education and the ways the different levels of decision making are to interrelate, three broad strategies could prove useful in improving the consistency of evaluation and assessment practices: tools and guidelines provided from the centre; collaboration among the different levels of authority, including partnerships for instance among school maintainers; and mechanisms to identify and share best practices within the education system.

A strategy involves initiatives at the central level to build up a knowledge base, tools and guidelines to assist evaluation and assessment activities. These typically include detailed plans to implement student learning objectives, including guidelines for schools and teachers to develop student assessment criteria. Examples of areas in which guidance from the centre could be useful are scoring guides and exemplars of different student performance levels teachers could use in their assessments; tools for teachers to use in the assessment of their students (e.g. test items banks); Internet platforms proposing formative teaching and learning strategies; tools for the self-appraisal of teachers; instruments for school leaders to undertake teacher appraisal; and tools and guidelines for school governing bodies to undertake the appraisal of school leaders.

Another strategy consists of encouraging collaboration between levels of authority within the system. For instance, in more decentralised countries, municipal partnerships could be encouraged to develop evaluation capacity, especially among the smallest municipalities. Another possibility is to promote the networking among the national, regional, municipal and private sector staff responsible for quality assurance in education. This could be done, for example, through an annual meeting of quality assurance staff at the different levels. The national and regional levels could also pay a greater role in supporting networks of municipalities (or groups of schools) working on particular quality assurance and improvement projects.

A further strategy involves benefitting to a higher degree from practice-based expertise and from the innovative practices developed at the local level. The national evaluation agencies and education authorities could play a greater role in disseminating and sharing effective practice across schools and local authorities. School governing bodies should be encouraged to collect examples of good practice from their schools. Evaluation agencies and national education authorities could provide guidance on how to select good examples, facilitate quality assurance of such examples, and feed evidence back to the system.

Use of results

Maintain sound knowledge management within the overall evaluation and assessment framework

Evaluation and assessment frameworks place great emphasis on the production of data and information on the results they create and their subsequent use for public information, policy planning and the improvement of practices across the system. This should be accompanied by sustained efforts to develop coherent information management systems to make the best use of the evidence generated by evaluation and assessment procedures across the system. Such systems would involve the establishment of protocols to harmonise, standardise, and share the data among key stakeholders.

An option is the development of a School/Education Portal, which is instrumental in ensuring access for stakeholders to a wide range of information about education, including results of evaluations. The Portal is typically a web-based information tool presenting key education monitoring information including learning outcomes, learning environments, resources and basic school data. The Portal could have an open part accessible to the general public and password-protected parts where evaluation agencies, schools, school leaders, teachers, parents and students can access information tailored to their needs.

Another option is to develop sophisticated data information systems – collection of data on students, teachers, schools, and their performance over time. These have the potential to assist teachers in the instruction of their students, provide quick feedback to school agents, serve as a platform to post relevant instructional material to support teachers and improve knowledge management, operate as a network to connect teachers and schools with similar concerns, and create a better data infrastructure for educational research. In addition, data management systems for schools to track progress of individual students should also be encouraged. These would ensure that such information can be shared among teachers or with a student's next school.

Commit to the use of evidence for policy development

In OECD countries there is a growing understanding of the importance of informing policies and the evaluation and assessment framework with evidence from research. Similarly, the rationale to establish an evaluation and assessment framework builds on the principle of using the results of evaluation and assessment to improve the knowledge base on which policy makers and practitioners draw to improve their practices. This calls for a strategic approach to research, analysis and evaluation, and information management activities in view of supporting the provision of evidence-based policy advice. Education authorities should promote a variety of research studies and analyses based on results from evaluation and assessment activities. This includes developing evaluation and assessment policies which are evidence-based.

Implementation

Anticipate potential implementation difficulties

The implementation of evaluation and assessment procedures requires the recognition of a range of important aspects. First, reaching agreements on the design of evaluation and assessment activities requires time for discussions and consultations with all stakeholders. Second, developing expertise in the system, including training evaluators is expensive and

requires time. Third, conducting evaluation processes induces additional workload for school agents. Fourth, aligning broader school reforms such as professional development opportunities with evaluation and assessment strategies requires more educational resources. It needs to be borne in mind that evaluation and the resulting feedback, reflection and development processes will only work if school agents make it work. To a great extent it is the motivated school agent who ensures the successful implementation of reforms in schools. Hence, it is imperative not only to find ways for school agents to identify with the goals and values of evaluation and assessment practices but also to ensure that such goals and values take account of school leader and teacher agency (OECD, 2006).

Engage stakeholders and practitioners in the design and implementation of evaluation and assessment policies

A range of strategies to consolidate the implementation of evaluation and assessment policies are available. To start with, the policy development process is more likely to yield consensus and compromise among parties if policies are developed through co-operation of different stakeholders towards a common goal. Indeed, regular interactions contribute, over time, to building trust among different stakeholders and raising awareness for the major concerns of others, thereby enhancing the inclination of the different parties for compromise. Educational evaluation policy has much more to gain from the cross-fertilisation of the distinct perspectives into compromises than from their antagonism and the imposition of one's views over other stakeholder groups. For instance, teachers will accept more easily to be evaluated if they are consulted in the design of the process. In addition to taking their fears and claims into account, the participation of teachers recognises their professionalism, the scarcity of their skills, and the extent of their responsibilities. If teacher appraisal procedures are unilaterally designed at the level of the administrative structure, without addressing and including the core of teaching practice, then there will be a "loose coupling" between administrators and teachers, that will both fail to provide public guarantees of quality, and will discourage reflection and review among teachers themselves (Elmore, 2000; Kleinhenz and Ingvarson, 2004). In more general terms, this calls for practitioners such as school leaders and teachers to be engaged in the design, management and analysis of evaluation and assessment policies. Consensus building among stakeholders is all the more important since local actors may be in the best position to foresee unintended consequences and judge what is feasible in practice.

Communicate the rationale for reform

Another priority is to clearly communicate a long-term vision of what is to be accomplished for student learning as the rationale for proposed evaluation and assessment policies. Individuals and groups are more likely to accept changes that are not necessarily in their own best interests if they understand the reasons for these changes and can see the role they should play within the broad national strategy. This includes dissemination of the evidence basis underlying the policy diagnosis, research findings on alternative policy options and their likely impact, as well as information on the costs of reform vs. inaction. Such communication and dissemination is critical to gain the support of society at large for educational evaluation reforms, not just the stakeholders with a direct interest.

Use pilots before full implementation and review implementation

Policy experimentation and the use of pilots may also prove effective strategies to overcome blockages dictated by disagreements among stakeholders and to assess the

effectiveness of policy innovations before generalising them. Policy makers also need to ensure mechanisms and platforms for the ongoing review and development of evaluation and appraisal systems to ensure they are up-to-date with latest research and developments (e.g. through advisory or steering groups).

In the same way, education practitioners should be provided opportunities to express their views and concerns on given evaluation and assessment initiatives as these are implemented. Implementation should involve feedback loops that allow adjustments to be made. School agents should be provided with opportunities to express their perceptions and concerns on evaluation processes as they are implemented. Interviews and surveys are common methods used to collect feedback on evaluation processes. The items generally include the understanding of the process, the acceptance of the standards, the fairness of the process and of the results, the capability and objectivity of the evaluators, the quality of the feedback received, the perceived impact of the evaluation process on practices and the overall impression of the evaluation system.

Ensure adequate capacity and sufficient resources

Furthermore, it is essential to develop capacity among stakeholders to implement evaluation and assessment policies. This includes providing support for school agents to understand evaluation procedures, training for evaluators to effectively undertake their responsibilities and preparation for school agents to use the results of evaluation. Evaluation and assessment are beneficial for improvement of educational practices provided that they engage the skills and commitment of practitioners.

Finally, there is a need for reducing excessive bureaucratic demands on schools and ensuring sufficient resources are provided in the implementation of evaluation and assessment policies. A consequence is that both those being evaluated and evaluators should be partly released from other duties. Schools agents should have time to reflect on their own practices, especially when the process requires self-appraisal and the constitution of a portfolio. Another aim should be reducing the administrative workload for evaluators, especially school leaders, in order to provide them with more time for evaluation activities, feedback and coaching.

Notes

1. As of 2013, Hungary experienced a trend towards a larger degree of central decision-making in education. See endnote 1 in Chapter 2.
2. TALIS is the OECD's Teaching and Learning International Survey, which was implemented in 2007-08, covering lower secondary education and with the participation of 23 countries (OECD, 2009). The results derived from TALIS are based on self-reports from teachers and principals and therefore represent their opinions, perceptions, beliefs and their accounts of their activities. Further information is available at *www.oecd.org/edu/talis*. The second cycle of TALIS (TALIS 2013) is being conducted in 2012-13.
3. Due to new regulations in Hungary on the role of education government in professional corporative bodies the Council for the Evaluation of Public Education ceased to exist legally in 2012.

References

ACARA (Australian Curriculum, Assessment and Reporting Authority) (2010), *Measurement Framework for Schooling in Australia 2010*, Sydney, Australia, *www.acara.edu.au/verve/_resources/Measurement_Framework_for_Schooling_in_Australia_2010.pdf*.

ADQS (Agency for the Development of Quality in Schools) (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Luxembourg*, Ministry of National Education and Vocational Training, Luxembourg, *www.oecd.org/edu/evaluationpolicy*.

Allerup, P., V. Kovac, G. Kvåle, G. Langfeldt and P. Skov (2009), *Evaluering av det nasjonale kvalitetsvurderingssystemet for grunnopplæringen,* Agderforskning, Kristiansand, Norway.

Alliance for Excellent Education (AEE) (2010), *Principles for a Comprehensive Assessment System*, Policy Brief.

Amsterdam, C.E., R.L. Johnson, D.E. Monrad and S.L. Tonnsen (2003), "A collaborative approach to the development and validation of a principal evaluation system: A case study", *Journal of Personnel Evaluation in Education,* Vol. 17, No. 3, pp. 221–242.

Avalos, B. and J. Assael (2006) "Moving from resistance to agreement: The case of the Chilean teacher performance evaluation", *International Journal of Educational Research*, Vol. 45, No. 4-5, pp. 254-266.

Baker, E. (2004), *Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Ballard, K. and A. Bates (2008), "Making a connection between student achievement, teacher accountability, and quality classroom instruction", *The Qualitative Report*, Vol. 13, No. 4, pp. 560–580.

Blondin, C. and B. Giot (2011), *Cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude thématique de l'OCDE : Rapport de base pour la Communauté française de Belgique*, Ministère de la Fédération Wallonie-Bruxelles, *www.oecd.org/edu/evaluationpolicy*.

Bransford, J.D., A.L. Brown and R.R. Cocking (1999), *How People Learn: Brain, Mind, Experience, and School*, National Academy Press, Washington, D.C.

Brejc, M., M. Sardoč and D. Zupanc (2011), *Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Slovenia,* National School for Leadership in Education in cooperation with the National Examinations Centre, *www.oecd.org/edu/evaluationpolicy*.

Brophy, J. (2006), *Grade Repetition*, The International Institute for Educational Planning (IIEP) and The International Academy of Education (IAE).

Campbell, C. and B. Levin (2008), "Using data to support educational improvement", *Educational Assessment, Evaluation and Accountability*, No. 21, pp. 47-65.

Chapman, J. and D. Aspin (2003), "Networks of learning: A new construct for educational provision and a new strategy for reform", in B. Davies and J. West-Burnham (eds.), *Handbook of Educational Leadership and Management*, Pearson, London, pp. 653-659.

Chudowsky, N. and J.W. Pellegrino (2003), "Large-scale assessments that support learning: What will it take?", *Theory into Practice*, Vol. 42, pp. 75-83.

Clarke, M. (2012), "What matters most for student assessment systems: A framework paper", *SABER – Systems Approach for Better Education Results, Working Paper No. 1*, The World Bank, Washington, DC.

Clifford, M. and S. Ross (2011), *Designing Principal Evaluation Systems: Research to Guide Decision-Making: An Executive Summary of Current Research*, paper prepared in collaboration with the National Association of Elementary School Principals, American Institutes for Research, Washington DC, *www.naesp.org/sites/default/files/PrincipalEvaluation_ExecutiveSummary.pdf*.

Commission on Instructionally Supportive Assessment (2001), *Building Tests to Support Instruction and Accountability*.

Corrales, J. (1999), "The politics of education reform: Bolstering the supply and demand, overcoming institutional blocks", *Education Reform and Management Series*, Vol. 2, No. 1.

Cullen, K. (1997), "An evaluation of the United Kingdom's national system of headteacher appraisal", *Studies in Educational Evaluation*, Vol. 23, No. 2, pp. 103-130.

Danielson, C. and T. McGreal (2000) *Teacher Evaluation to Enhance Professional Practice*, Association for Supervision and Curriculum Development (ASCD), Alexandria, Virginia.

Danish Ministry of Education and Rambøll (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Denmark, www.oecd.org/edu/evaluationpolicy*.

Datnow, A. and M. Castellano (2000), "Teachers' responses to success for all: How beliefs, experiences and adaptation shape implementation", *American Educational Research Journal*, Vol. 37, No. 3, pp. 775-799.

Davis, S.H. and P.A. Hensley (1999), "The politics of principal evaluation", *Journal of Personnel Evaluation in Education*, Vol. 13, No. 4, pp. 383-403.

DEEWR (Department of Education, Employment and Workplace Relations) (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Australia*, Canberra, *www.oecd.org/edu/evaluationpolicy*.

Dos Santos, S. and R. Rakocevic (2012), *Les cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude de l'OCDE : Rapport national de base pour la France*, Direction de l'Évaluation, de la Prospective et de la Performance (DEPP), Ministère de l'Éducation nationale, *www.oecd.org/edu/evaluationpolicy*.

Eivers, E., S. Close, G. Shiel, D. Millar, A. Clerkin, L. Gilleece and J. Kiniry (2010), *The 2009 National Assessment of Mathematics and English Reading in Primary Schools, www.erc.ie/documents/na2009_report.pdf.*

Elmore, R. (2000), *Building a New Structure for School Leadership*, Albert Shanker Institute, Winter 2000.

Elmore, R. and D. Burney (1997), *Investing in Teacher Learning: Staff Development and Instructional Improvement in Community School District #2*, National Commission on Teaching & America's Future & the Consortium for Policy Research in Education, New York.

Fazekas, M. and T. Burns (2012), "Exploring the complex interaction between governance and knowledge in education", *OECD Education Working Papers*, No. 67, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Field, S., M. Kuczera and B. Pont (2007), *No More Failures: Ten Steps to Equity in Education*, OECD Publishing, Paris.

Finlay, I., S. Niven and S. Young (1998), "Stakeholders, consensus, participation and democracy", in I. Finlay, S. Niven and S. Young (eds.), *Changing Vocational Education and Training: An International Comparative Perspective*, Routledge, London.

Finn, C.E. and M. Kanstoroom (2001), "State academic standards", *Brookings Papers on Education Policy*, No. 4, pp. 131-164.

Finnish Ministry of Education and Culture (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Finland*, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Fiske, E. (1996), *Decentralization of Education – Politics and Consensus*, World Bank, Washington, D.C.

Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, *www.oecd.org/edu/evaluationpolicy*.

Fournier, G. and D. Mildon (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Canada*, Council of Ministers of Education, Canada (CMEC).

Gamoran, A. et al. (1997), "Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth", *Educational Evaluation and Policy Analysis*, Vol. 19, pp. 325–338.

Gordon, J., G. Halász, M. Krawczyk, T. Leney, A. Michel, D. Pepper, E. Putkiewicz and J. Wisniewski (2009), *Key Competences in Europe: Opening Doors for Lifelong Learners across the School Curriculum and Teacher Education* (2009), Study undertaken for the Directorate General Education and Culture of the European Commission, CASE-Center for Social and Economic Research, Warsaw, *http://ec.europa.eu/education/more-information/moreinformation139_ en.htm.*

Haertel, E. and J. Herman (2005), *A Historical Perspective on Validity Arguments for Accountability Testing*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Haertel, E.H. and W.A. Lorié (2004), "Validating standards-based test score interpretations", *Measurement: Interdisciplinary Research and Perspectives*, Vol. 2, pp. 61-103.

Hamilton, L. and B. Stecher (2002), "Improving test-based accountability", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education,* RAND Publishing, Santa Monica, California.

Hargreaves, D.H. (2003), *Education Epidemic: Transforming Schools through Innovation Networks*, Demos, London.

Heneman, H., A. Milanowski, S. Kimball and A. Odden (2006), "Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay", *Consortium for Policy Research in Education (CPRE) Policy Briefs,* RB-45.

Hess, F. and M. West (2006), *A Better Bargain: Overhauling Teacher Collective Bargaining for the 21st Century*, Program on Education Policy and Governance, Harvard University, Cambridge, Massachusetts.

Hill, P., C. Campbell and J. Harvey (2000), *It Takes a City: Getting Serious about Urban School Reform*, The Brookings Institution, Washington, D.C.

Hooge, E., T. Burns and H. Wilkoszewski (2012), "Looking Beyond the Numbers: Stakeholders and Multiple School Accountability", *OECD Education Working Papers*, No. 85, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Hooge, E.H. and J.K. Helderman (2008). "Klant en overheid koning. Over toezicht op meervoudige verantwoording door maatschappelijke ondernemingen" ["Client and Government King? About Monitoring Internally Horizontal Accountability by Dutch Societal Entrepreneurs"], *Bestuurskunde*, 17(3), pp. 95-104.

Hungarian Ministry of Education and Culture (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Hungary*, *www.oecd.org/edu/evaluationpolicy*.

IIE (Institute for Information on Education) (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Czech Republic*, *www.oecd.org/edu/evaluationpolicy*.

Ingvarson, L., E. Kleinhenz and J. Wilkinson (2007), *Research on Performance Pay for Teachers*, Australian Council for Educational Research (ACER).

INVALSI (2010), *Sistema di Audit e Accompagnamento delle Scuole PON nella prospettiva dell'Apprendere ad Apprendere – Rapporto Fase 1 Pilota,* by R.M. Frascati, *http://archivio.pubblica.istruzione.it/fondistrutturali/documenti/valutazione2007.shtml*.

Irish Department of Education and Skills (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Ireland, www.oecd.org/edu/evaluationpolicy*.

Katz, S., L.M. Earl and S. Ben Jaafar (2009), *Building and Connecting Learning Communities: The Power of Networks for School Improvement*, Corwin, Thousand Oaks, California.

Keating, B. (2011), *How Four Districts Crafted Innovative Principal Evaluation Systems: Success Stories in Collaboration*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1109.pdf*.

Kennedy, M. (2005) *Inside Teaching*, Harvard University Press, London, England.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and Jungha Park (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

Kleinhenz, E. and L. Ingvarson (2004), "Teacher evaluation uncoupled: A discussion of teacher evaluation policies and practices in Australian states and their relation to quality teaching and learning", *Research Papers in Education*, Vol. 19, No. 1, pp. 31-49.

Koretz, D. (2005), *Alignment, High Stakes, and the Inflation of Test Scores*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Le, V. and S. Klein (2002), "Technical criteria for evaluating tests", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education,* RAND Publishing, Santa Monica, California.

Leithwood, K. (2001), "School leadership in the context of accountability policies". *Leadership in Education*, 4(3), pp. 217-235.

Leithwood, K., R. Steinbach and D. Jantzi (2002), "School leadership and teachers' motivation to implement accountability policies", *Educational administration Quarterly*, Vol. 38, No. 1, pp. 94-119.

Leon, R., S. Davis, K. Kearney, N. Sanders and C. Thomas (2011), *Key Features of a Comprehensive Principal Evaluation System*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1107.pdf*.

Linn, R.L. (2005), *Issues in the Design of Accountability Systems*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Linn, R.L. (2001), *The Design and Evaluation of Educational Assessment and Accountability Systems*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Linn, R. (1998), *Assessments and Accountability*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Little, J.W. (1990), "The persistence of privacy: Autonomy and initiative in teachers' professional relations", *Teachers College Record,* Vol. 91, pp. 509–536.

Looney, J. (2011a), "Alignment in complex systems, achieving balance and coherence", *OECD Education Working Papers*, No. 64, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Looney, J. (2011b), "Integrating formative and summative assessment: Progress toward a seamless system?", *OECD Education Working Papers*, No. 58, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Looney, J. (2009), "Assessment and innovation in education", *OECD Education Working Papers*, No. 24, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Marshall, K. (2005), "It's time to rethink teacher supervision and evaluation", *Phi Delta Kappan*, Vol. 86, No. 10, pp 727-735.

MCEECDYA (2009), *Principles and Protocols for Reporting on Schooling in Australia*, *www.mceecdya.edu.au/mceecdya/nap_principles__protocols_for_rep_on_school_2009,27896.html*.

McLaughlin, M.W. and J.D. Talbert (2001), *Professional Communities and the Work of High School Teaching,* University of Chicago Press, Chicago.

Milanowski, A. and H. Heneman (2001), "Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study", *Journal of Personnel Evaluation in Education*, Vol. 15, No. 3, pp. 193-212.

MIUR (Ministry of Education, University and Research of Italy) (2011), *La scuola in cifre 2009-10*, Rome, *www.istruzione.it/alfresco/d/d/workspace/SpacesStore/08c4b54d-18e4-497c-be1b-5bec9927e388/scuolaincifre_2009-2010.pdf*.

Mons, N. (2009), "Theoretical and real effects of standardised assessment", background paper to the study: National Testing of Pupils in Europe, Eurydice Network, *http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111EN.pdf*, accessed 12 July 2011.

Morris, A. (2011), "Student standardised testing: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 65, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Newman, F.M., B. Smith, E. Allensworth and A.S. Bryk (2001), "Instructional program coherence: What it is and why it should guide school improvement policy", *Educational Evaluation and Policy Analysis,* Vol. 23, No. 4, pp. 297-321.

Newman, F.M. and G.G. Wehlage (1995), *Successful School Restructuring: A Report to the Public and Educators by the Center on Organization and Restructuring of Schools*, Center on Organization and Restructuring of Schools, Madison, Wisconsin.

New Zealand Ministry of Education (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for New Zealand, www.oecd.org/edu/evaluationpolicy*.

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

O'Day, J. (2002), "Complexity, accountability and school improvement", *Harvard Education Review*, Vol. 72, No. 3.

OECD (2012a), *Education at a Glance: OECD Indicators 2012*, OECD Publishing, Paris.

OECD (2012b), *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing, Paris.

OECD (2011), *Education at a Glance: OECD Indicators 2011*, OECD Publishing, Paris.

OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

OECD (2007), *Evidence in Education: Linking Research and Policy*, OECD Publishing, Paris.

OECD (2006), "The teaching workforce: Meeting aspirations and enhancing motivation", in *Education Policy Analysis 2005*, OECD Publishing, Paris.

Olsen, J.P. (1989), *Modernization Programs in Perspective – Institutional Analysis of Organizational Change*, Notat 89/46, LOS-Senter, Bergen.

Pont, B., D. Nusche, H. Moorman (2008), *Improving School Leadership Volume 1: Policy and Practice*, OECD Publishing, Paris.

Porter, A.C. and J.L. Smithson (2001), "Defining, developing and using curriculum indicators", *CPRE Research Report Series,* RR-048, Consortium for Policy Research in Education, Philadelphia.

PriceWaterhouseCoopers (PwC) (2007), *Independent Study into School Leadership: Main Report*, Department of Education and Skills, London, England.

Rhode Island Department of Elementary and Secondary Education (2009), *Education Evaluation System Standards*, Providence, Rhode Island, *www.ride.ri.gov/regents/Docs/RegentsRegulations/RI%20Educator%20Evaluation%20Standards%208-06-09%20Public%20Comment%20Version.pdf*.

Rosenkvist, M. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Santiago, P., F. Benavides, C. Danielson, L. Goe and D. Nusche (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012b), *OECD Reviews of Evaluation and Assessment in Education: Mexico*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012c), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., D. Roseveare, G. van Amelsvoort, J. Manzi and P. Matthews (2009), *Teacher Evaluation in Portugal: OECD Review*, OECD, Paris, *www.oecd.org/edu/teacherevaluationportugal*.

Sartain, L., S.R. Stoelinga and E.R. Brown; with S. Luppescu, K.K. Matsko, F.K. Miller, C.E. Durwood, J.Y. Jiang and D. Glazer (2011), *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation*, Consortium on Chicago School Research at the University of Chicago, Urban Education Institute, Research Report, Chicago, *http://ccsr.uchicago.edu/sites/default/files/publications/Teacher%20Eval%20Report%20FINAL.pdf*.

Sawyer, R. K. (2008), "Optimising Learning: Implications of Learning Sciences Research", *Innovating to Learn, Learning to Innovate*, OECD Publishing, Paris.

Scheerens, J., M. Ehren, P. Sleegers and R. de Leeuw (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Netherlands,* University of Twente, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Schmidt, W.H. and A. Maier (2009), "Opportunity to learn", in G. Sykes, B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, AERA, London and New York, pp. 541-559.

Seashore Louis, K., K. Leithwood, K.L. Wahlstrom and S.E. Anderson (2010), *Investigating the Links to Improved Student Learning: Final Report of Research Findings"*, University of Minnesota, St. Paul, *www.cehd.umn.edu/CAREI.*

SEP and INEE (Secretariat of Public Education and National Institute for Educational Assessment and Evaluation (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Mexico*, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., J. van Bruggen, D. Nusche and P. Wright (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Slovak Republic*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., M. Hulshof, D. Nusche and L. Stoll (2011a), *OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011b), *OECD Reviews of Evaluation and Assessment in Education: Denmark,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Sliwka, A. (2008), "The contribution of alternative education", *Innovating to Learn, Learning to Innovate*, OECD Publishing, Paris.

Smithson, J.L. and A.C. Collares (2007), "Alignment as a predictor of student achievement gains", paper presented at the annual meeting of the American Educational Research Association, Chicago, April.

Specht, W. and F. Sobanski (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Austria*, Bundesminiterium für Unterricht, Kunst und Kultur (BMUKK), Vienna, *www.oecd.org/edu/evaluationpolicy*.

Stecher, B.M., S. Barron, T. Chun and K.E. Ross (2000), *The Effects of the Washington State Education Reform on Schools and Classrooms: Initial Findings*, DB-309-EDU, RAND Santa Monica, California.

Thomas, D.W., E.A. Holdaway and K.L. Ward (2000), "Policies and Practices Involved in the Evaluation of School Principals", *Journal of Personnel Evaluation in Education*, Vol. 14, No. 3, pp. 215-240.

Wurzburg, G. (2010), "Making reform happen in education", in *Making Reform Happen: Lessons from OECD Countries*, OECD Publishing, Paris.

Xia, N. and S.N. Kirby (2009), *Retaining Students in Grade: A Literature Review of the Effects of Retention on Students' Academic and Nonacademic Outcomes*, prepared for the New York City Department of Education, RAND Corporation, Santa Monica.

Chapter 4

Student assessment:

Putting the learner at the centre

Student assessment is essential to measure the progress and performance of individual students, plan further steps for the improvement of teaching and learning, and share information with relevant stakeholders. This chapter describes the approaches that countries take to assess individual students. Building on a discussion of impact, drivers and contextual developments, it discusses the governance of student assessment systems, assessment procedures and instruments, capacities needed for effective student assessment and the use of assessment results for different purposes. The chapter concludes with a set of pointers for policy development.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter considers approaches to student assessment within the evaluation and assessment framework. It focuses on how student assessment influences the learning experience of individual students and considers both summative assessment (assessment *of* learning) and formative assessment (assessment *for* learning) of students. The chapter does not cover the use of aggregated student assessment results to make judgements about the performance of teachers, schools and education systems, because these issues will be addressed in the following chapters.[1]

Assessment is a process that helps focus attention towards what matters most in education, beyond just access and participation: the actual learning outcomes of each student. Gathering information on where students stand in their learning and the progress that they have made is key to designing strategies for the further improvement of teaching and learning. Sharing such information with stakeholders across the education system is essential to meet information needs and support decision making at the classroom, school and education system level.

This chapter is organised in eight sections. After this introduction, the second section presents the analytical approach, followed by a third section on impact, drivers and contextual developments. The following four sections describe key features of student assessment and country practices, structured around the main topics of the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: governance, procedures, capacity and use of results. The final section provides pointers for policy development.

Analytical approach

Definitions

Student assessment refers to processes in which evidence of learning is collected in a planned and systematic way in order to make a judgement about student learning (EPPI, 2002). It encompasses summative and formative purposes, and may be designed and implemented internally within the school or externally through standardised assessments.

Summative and formative assessment

The assessment literature has traditionally made a distinction between assessment for summative purposes and assessment for formative purposes. Some authors also make a distinction between formative assessment and diagnostic assessment, but throughout this report diagnostic assessment will be considered as one aspect of formative assessment.

- **Student summative assessment**, or assessment *of* learning, aims to summarise learning that has taken place, in order to record, mark or certify achievements (EPPI, 2002).
- **Student formative assessment**, or assessment *for* learning, aims to identify aspects of learning as it is developing in order to deepen and shape subsequent learning.
- **Diagnostic assessment** is one type of formative assessment, which often takes place at the beginning of a study unit in order to find a starting point, or baseline, for learning and to develop a suitable learning programme. Diagnostic assessment

may also serve to identify students who are at risk of failure, to uncover the sources of their learning difficulties and to plan for an appropriate supplemental intervention or remediation.

In practice, the purposes for different assessment approaches are not always clearly stated and results from the same assessment processes may be used for either summative or formative purposes. How to establish the right balance between summative and formative assessment and how to achieve each purpose most effectively with different assessment formats are key questions considered throughout this chapter. The chapter will also consider if, and to what extent, summative and formative assessment can be integrated effectively.

Internal and external assessment

Another important distinction has traditionally been made between internal (school-based) assessment and external (standardised) assessment. It is important to note that both internal and external assessments may be used in a summative or formative way.

- **Internal assessment**, **or school-based assessment**, is designed and marked by the students' own teachers, often in collaboration with the students themselves, and implemented as part of regular classroom instruction, within lessons or at the end of a teaching unit, year level or educational cycle.
- **External assessment**, **or standardised assessment**, is designed and marked outside individual schools so as to ensure that the questions, conditions for administering, scoring procedures, and interpretations are consistent and comparable among students (Popham, 1991). External assessments may be applied to a full student cohort or only in some schools and classrooms (for example, on-demand assessments that schools can use to measure their own progress and benchmark themselves against national averages).

In practice, however, the distinctions between internal and external assessments are not always so clear-cut. For example, there are also hybrid forms of assessment that are developed externally but implemented and marked internally by the students' own teachers. How to best design assessment frameworks drawing on a mix of internal and external approaches in order to achieve stated summative and formative purposes will be another guiding question throughout this chapter.

Key concepts related to student assessment

Designing assessments in a way that they are fit for the intended purpose is important to ensure their reliability, validity, transparency and usability. These terms are briefly defined below, as they will be used frequently throughout the chapter.

- **Validity** relates to the appropriateness of the inferences, uses and consequences attached to assessment. A highly valid assessment ensures that all relevant aspects of student performance are covered by the assessment.
- **Reliability** refers to the extent to which the assessment is consistent in measuring what it sets out to measure. A highly reliable assessment ensures that the assessment is accurate and not influenced by the particular assessor or assessment occasion.

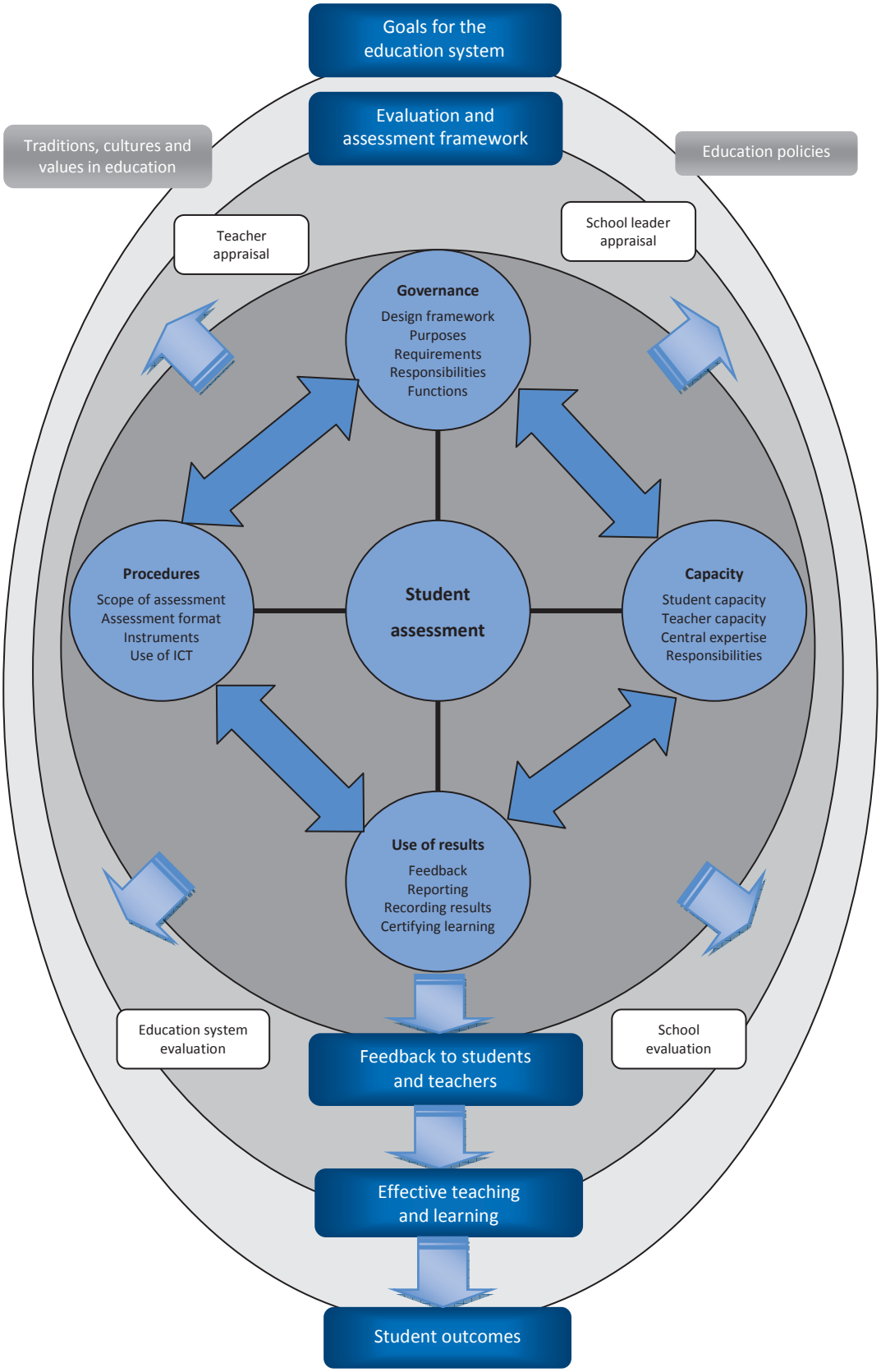
- **Transparency** relates to the degree to which information is available regarding expected learning outcomes, the criteria that will be used in judging student learning and the rules being applied when judgements are made about learning.
- **Usability** refers to how policy makers, school leaders, teachers, parents and students make sense of and respond to assessment results. An objective in designing student assessment is to maximise the value of the assessment by making it timely, easy to understand and interpret for teachers and/or students, and instructionally useful, at the right grain size to guide subsequent, intended decision making and action.

Conceptual framework

This chapter aims to explore the complex range of features associated with student assessment. Figure 4.1 below provides a conceptual framework summarising the aspects involved and the way they interconnect. The overarching policy objective is to ensure that student assessment contributes to the improvement of student outcomes through improved teaching and learning. The conceptual framework has four main interrelated themes.

- **Governance**: This first section deals with the governance of student assessment systems across OECD countries. It describes the different purposes and objectives of student assessment systems and the legal frameworks in place to ensure that student assessment results are used in a way that such objectives are reached. The section also explores how responsibilities for assessment are distributed in different countries and how different levels of governance interact to form a coherent assessment system.
- **Procedures**: This second section describes the procedures and methodologies used for student assessment across countries. This includes the scope of assessment, i.e. the areas of learning that are covered by the assessment as well as the key procedural features of student assessment across countries, i.e. the mix of instruments used in specific student assessment systems; the format of assessments; and the use of ICT in assessment. It also reviews ways in which the design of assessments can enhance or threaten fairness and equity in education.
- **Capacity**: This third section discusses the competencies and the support necessary to assess students, to benefit from assessment, and to use the results of student assessment. It includes issues such as: the capacities students need to engage in and learn from their assessment; the assessment competencies that teachers acquire in initial teacher education, professional development and moderation arrangements; and the expertise of the agencies involved in student assessment.
- **Use of results**: This fourth section is concerned with how assessment results are reported and used for both summative and formative purposes. It describes standards of quality and reporting formats used in different contexts, reviews the legal frameworks in place to regulate reporting of results and discusses the ways in which assessment results are used in different contexts to record information, provide feedback to students and make decisions about their further educational trajectory.

Figure 4.1 Conceptual framework for student assessment



Impact, drivers and contextual developments

The importance and impact of student assessment

Assessment helps focus attention on the learning progress and outcomes of each student. Collecting student assessment information is essential to improve teaching and learning strategies and meet information needs at the level of students, parents, teachers, school leaders, policy makers and the general public.

Students need to be clear about what they are aiming to learn and which indicators and criteria are appropriate to evaluate progress and inform future learning. Engaging students as active participants in assessment will help them develop capabilities in analysing their own learning and becoming self-directed learners. *Parents* typically want to know how their children are doing and progressing in relation to expected standards and in comparison to others in the same age group. Providing assessment information to parents is key to building strong school-home partnerships by making parents aware of learning goals, their children's progress and priorities for further learning.

Teachers need assessment information that is reliable and consistent across schools in order to understand student strengths and weaknesses in relation to expected standards, to target future teaching and improve classroom instruction. *School leaders* can use such information for school self-evaluation processes and to provide accountability information to their employers and the educational administration (Chapter 6). *Policy makers* need aggregated assessment information to monitor the performance of schools and education systems and ensure that national education goals are met (Chapter 8). *Society* at large also needs credentials about the quality of education and the achievement of standards in the education system (Chapter 8).

There is a large body of research showing a strong impact of different types of assessment on student learning outcomes (Box 4.1). Evidence on different approaches indicates that assessment may support or diminish student motivation and performance depending on the way it is designed, implemented and used. In other words, assessments that are not well designed and implemented may in fact contribute to alienating students (and teachers) from the education system and exacerbate inequity in education. On the other hand, carefully planned assessment interventions that are well aligned with learning goals and place students at the centre of the process have strong potential to raise achievement and reduce disparities.

Drivers and contextual developments

Before moving to the analysis of key features of assessment, this section aims to set the context in which student assessment takes place across OECD countries. Student assessment, like all components of evaluation and assessment frameworks, is influenced by wider trends and developments shaping education policies (see Chapter 2). New understandings of the nature and purpose of learning and assessment have shaped assessment policies in all countries. This section provides a brief overview of the key contextual developments impacting on student assessment policy and practice.

Box 4.1 How student assessment influences learning outcomes: A brief overview of research evidence

Empirical research on the impact of education policies and practices on student learning outcomes is conceptually and methodologically challenging. Learning outcomes are shaped by a range of extra- and intra-institutional factors including family background, abilities and attitudes, organisation and delivery of teaching, school practices and characteristics of the education system. Studies measuring the impact of different education policies on student achievement tend to use data sets and methodologies providing limited measures of learning and partial indicators of the range of important factors. The outcomes and policy recommendations of such research are sometimes contested, especially when they tend to generalise results across different contexts.

Bearing these limitations in mind, a range of policy-relevant conclusions can nonetheless be drawn from the numerous studies exploring the link between student assessment approaches and learning outcomes. This brief overview of research draws on large-scale quantitative studies, experimental studies and case study evaluations. Given the sheer number of relevant studies, specific targeted searches were also made for prior reviews and meta-analyses regarding the impact of different assessment approaches. These helped to uncover different conceptual and methodological strands of the literature and to make sense of contradictory research findings.

Formative classroom assessment

A large amount of research has been conducted around the world regarding the impact of formative assessment on learning outcomes. In their seminal review of the research on classroom-based formative assessment, Black and Wiliam (1998) brought together evidence gathered from 250 international sources regarding the use and impact of formative assessment. The 250 sources reviewed for this purpose cover learners ranging pre-school to university. Evidence of impact was drawn from more than 40 studies conducted under ecologically valid circumstances (that is, controlled experiments conducted in the student's usual classroom setting and with their usual teacher). They included studies on effective feedback; questioning; comprehensive approaches to teaching and learning featuring formative assessment, and student self- and peer-assessment. Black and Wiliam concluded that the achievement gains associated with formative assessment were among the largest ever reported for educational interventions. The review also found that formative assessment methods were, in some cases, particularly effective for lower achieving students, thus reducing inequity of student outcomes and raising overall achievement. The 1998 Black and Wiliam review confirmed earlier reviews by Natriello (1987) and Crooks (1988), which had reached substantially the same conclusions (Looney, 2011a).

At the same time, the success of formative assessment policies depends very much on their effective implementation (Black, 1993; Black and Wiliam, 1998; Stiggins et al., 1989). The quality of formative assessment rests, in part, on strategies teachers use to elicit evidence of student learning related to goals, with the appropriate level of detail to shape subsequent instruction (Bell and Cowie, 2001; Heritage, 2010; Herman et al., 2010). But in some contexts, it is still more typical for teachers to develop only superficial questions to probe student learning, and provide only general feedback (Swaffield, 2008). Teachers may have difficulty in interpreting student responses or in formulating next steps for instruction (Herman et al., 2010). And while many teachers agree that formative assessment methods are an important element in high-quality teaching, they may also find that there are logistical barriers to making formative assessment a regular part of their teaching practice, such as large classes, extensive curriculum requirements, and the difficulty of meeting diverse and challenging student needs (OECD, 2005a; Looney, 2011a). This highlights the importance of firmly embedding formative assessment within the broader evaluation and assessment framework and the need to support teachers' capacity and professionalism in formative assessment.

Summative classroom assessment

Summative classroom assessment activities are a substantial part of education across OECD countries. The strong impact of summative assessment on teaching and learning has been widely reported. In many contexts, summative assessment dominates what students are oriented towards in their learning – this is typically described as the "backwash effect" of summative assessment (Alderson and Wall, 1993; Somerset, 1996; Biggs, 1998; Baartman et al., 2006). The use of summative assessment often rests on the assumption that if the assessment matters to students they will seek to influence the result by increasing effort and improving performance (Becker and Rosen, 1992).

Box 4.1 How student assessment influences learning outcomes: A brief overview of research evidence (*continued*)

Hence, the need to perform on a test or to hand in an important assignment may concentrate and energise students' learning activities. The marks, transcripts and diplomas that summarise student performance can be seen as rewards for student effort and achievement, which provide an extrinsic motivation for learning (Sjögren, 2009). Some studies have shown that students who lack intrinsic motivation in a specific area in the first place can be stimulated to develop interest in the area via carefully planned experiences of extrinsic rewards (Crooks, 1988).

However, reviews of research in this field suggest that the use of extrinsic motivation may be problematic, because such extrinsic motivation is closely related to the reward (Crooks, 1988; EPPI, 2002). This means that where external rewards are provided, learning will be targeted to those domains that are rewarded, and that effort may decrease or disappear when the reward is no longer provided (Crooks, 1988; Kohn, 1994). There are risks that summative assessments with high stakes for students may in fact encourage surface learning approaches, generate ego-related priorities, reduce enjoyment of learning and decrease student focus on long-term goals (Biggs, 1998; EPPI, 2002). In the education context, studies repeatedly indicated that students with strong initial motivation might be negatively affected by attempts to stimulate their learning by external rewards (Crooks, 1988). Hidi and Harackiewicz (2000) question the dichotomy of intrinsic and extrinsic motivation. In their review of research on the role of interest and goals for achievement, they suggest that it may be necessary to combine intrinsic rewards (via activities that are inherently interesting to students) with external rewards in order to support optimal and sustained learning efforts. This points to the need to develop assessment frameworks where a range of formative and summative assessment approaches complement each other to provide the adequate level of challenge and support to each student.

External assessments and examinations

An extensive body of large-scale quantitative research deals with the effects of external exit examinations at the end of upper secondary education on student learning. Evidence from several empirical cross-country studies suggests that students in countries that have external exit examinations in place perform significantly better on international student assessments than students in countries that do not have such examinations (Bishop, 1997; 1999; 2006; Woessmann et al., 2009). These results are corroborated by a number of cross-regional studies conducted in the United States, Canada and Germany (Graham and Husted, 1993; Bishop, 1999; Luedemann, 2011). Some researchers have emphasised the strong role of external assessments in motivating teachers and students for achievement. Externally defined assessments can clearly indicate the standards that are expected nationally and signal to students and teachers what needs to be learned. This can be a way of making sure that high standards are expected of *all* students (Elshout-Mohr et al., 2002; Rawlins et al., 2005, Kellaghan et al., 1996). Assessment that is externally administrated can also positively influence teacher-student relationships, as the teacher becomes an ally of students in preparing the external examination rather than a judge (Bishop, 2006).

At the same time, several studies have found potential negative effects of external exit examinations. For students who feel the standards are set too high, exit examination may lead to loss of motivation and increased drop-out rates, especially so for low-income, minority and low-performing students (Clarke et al., 2000; Dee and Jacob, 2006; Papay et al., 2008; Ou, 2010). Hence, even if exit examinations may enhance overall student performance by clarifying expected learning and increasing student motivation, these positive effects may be mitigated by higher numbers of drop-outs and reduced opportunities for disadvantaged students (Greaney and Kellaghan, 1995; Dufaux, 2012). Also, where high stakes are attached to external assessments, distortions in the education process may occur, such as excessive focus on teaching students the specific skills that are assessed, narrowing the curriculum, distributing repeated practice tests, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and sometimes even outright manipulation of results (Koretz et al., 1991; Klein et al., 2000; Linn, 2000; Stecher and Barron, 2001; Clarke et al., 2003; Jacob, 2005; McMurrer, 2007; Hamilton et al., 2007; Sims, 2008; Stiggins, 1999; Slomp, 2008) (more on this in Chapter 6). Because of these potential negative effects, it is important to establish safeguards against excessive emphasis on a particular standardised test and to draw on a range of assessment information to make judgements about learning progress.

Towards a new understanding of learning

The national curricula in many OECD education systems have been reformed in recent years to emphasise the development of complex competencies rather than a narrow focus on isolated knowledge and skills. Typically, curricula in primary and secondary education now feature a list of key competencies that the education system should seek to promote across all subjects and year levels (Box 4.2). While the definitions of key competencies vary considerably across countries, they reflect a similar ambition: overcoming traditional educational approaches focussing primarily on knowledge transmission and acquisition of basic skills. The aim of many recent curriculum reforms is to promote a broader model of learning which comprises a complex integration of knowledge, skills, attitudes and action in order to carry out a task successfully in real-life contexts. Such key competencies, or "21st century skills", typically include dimensions such as critical thinking, creativity, problem-solving, communication, ICT literacy, as well as collaborative, social and citizen skills (Box 4.2).

An important similarity of most definitions of key competencies is a shared focus on "learning for life", "lifelong learning" or "learning to learn". While they emphasise different elements, these terms are clear in suggesting that what is learned must have relevance beyond school (Lucas and Claxton, 2009). This responds to a concern that school settings sometimes tend to promote a narrow set of cognitive skills and attitudes which have limited relevance outside the classroom, such as taking accurate handwritten notes, remembering detailed information acquired months or years ago and sitting still for long periods of the day. While these may be essential "school skills", they are insufficient to equip students for active participation in society and the world of work in the future (Lucas and Claxton, 2009).

In particular, the exponential increase in the availability of information has made it less important for learners to be able to recall and reproduce facts while making it more important to develop competencies to synthesise, transform and apply learning in real-world situations, think creatively and critically, collaborate with others, communicate effectively and adapt to rapidly developing environments. Hence, while numeric, verbal and scientific literacy will remain important building blocks of education, more generic and transversal competencies are becoming increasingly important (European Commission, 2011a, 2012).

As expectations of what students should achieve have changed, there has been parallel reflection on how to best design assessment approaches that can actually measure such broader competencies. Given the strong backwash effect of assessment on learning (Box 4.1), innovations in pedagogy are unlikely to be successful unless they are accompanied by related innovations in assessment (Cizek, 1997). For assessment to be meaningful, it must be well-aligned to the type of learning that is valued. For example, factual knowledge tests are well-suited to assess the outcomes of traditional teaching approaches based on rote learning and knowledge transfer. But such tests are less adequate when it comes to assessing complex competencies (Biggs, 1996, 1999).

Box 4.2 Key competencies around the world

In Europe, with the Recommendation on Key Competences (2006), all EU member states have agreed on a framework of eight "key competencies" that are seen as necessary for personal fulfilment and development, active citizenship, social inclusion and employment (European Commission, 2011a). These include competencies in communication, mathematics, science and technology as well as learning to learn, social and civic competencies, sense of initiative, entrepreneurship and cultural awareness and expression. As expressed in the European Framework, these competencies are underpinned by process dimensions such as critical thinking, creativity, problem solving and decision taking (European Commission, 2011b, 2012). The European focus on key competencies is reflected across national curricula in Europe, with most EU member states reporting that they have already changed their primary and secondary school curricula to incorporate elements of the key competences or even the complete framework.

Similar trends can be observed beyond Europe. The United Nations Educational, Scientific and Cultural Organisation (UNESCO) has been promoting a focus on "life skills", which it defines as including competencies such as critical thinking, creativity, ability to organise, social and communication skills, adaptability, problem solving, ability to co-operate on a democratic basis that are needed for actively shaping a peaceful future (Singh, 2003). The terminology of life skills has been included in several national curricula. In Mexico, for example, the curriculum for basic education was reformed in 2011 around five "competencies for life" as promoted by UNESCO: lifelong learning; information management; management of situations; coexistence; and life in society (Santiago et al., 2012c).

In the United States, the term most commonly used is that of "21st century skills" or "21st century competencies", which was defined by Binkley et al. (2010) as including (1) ways of thinking (creativity, innovation, critical thinking, problem solving, decision making, learning to learn, metacognition); (2) ways of working (communication, collaboration), (3) tools for working (information literacy, ICT literacy); and (3) living in the world (local and global citizenship, life and career, personal and social responsibility). According to the Partnership for 21st Century Skills, the focus on 21st century skills has now been incorporated into the educational systems of 16 states in the United States (*www.p21.org*). In Canada, also, all jurisdictions have, to a varying degree, reshaped curriculum from knowledge-based curriculum to performance-based curriculum, with a new emphasis on problem solving and cognitive application of knowledge using higher-level skills beyond recall and comprehension.

In Australia, the Melbourne Declaration on Educational Goals for Young Australians, released in December 2008, and agreed to by all education ministers through the Ministerial Council on Employment, Training and Youth Affairs (MCEETYA, 2008) sets the overarching goal that all young Australians become successful learners, confident and creative individuals, and active and informed citizens (Santiago et al., 2011). Before the introduction of the Australian Curriculum (which includes a set of General Capabilities encompassing the knowledge, skills, behaviours and dispositions that, together with curriculum content in each learning area, will assist students to live and work successfully in the 21st century), individual Australian states had already done much pioneering development of broader competency standards and frameworks.

The New Zealand Curriculum, revised in 2007, is organised around five key competencies: thinking, using language, symbols and text, managing self, relating to others, participating and contributing. The curriculum highlights that "people use these competencies to live, learn, work and contribute as active members of their communities. More complex than skills, the competencies draw also on knowledge, attitudes, and values in ways that lead to action. They are not separate or stand-alone. They are the key to learning in every area."

Across the OECD, several public and private actors are increasingly investing in research and development regarding the teaching and assessment of key competencies. One example is the Assessment and Teaching of 21st Century Skills (ATC21s) project at The University of Melbourne, Australia, which is sponsored by private companies (Cisco, Intel and Microsoft) and governed by an executive board comprising ministries of education, academics and industry leaders from a range of countries (*www.atc21s.org*).

Sources: European Commission (2011a, 2011b, 2012); Singh (2003); Santiago et al. (2011, 2012c); Fournier and Mildon (forthcoming); MCEETYA (2008); Partnership for 21st Century Skills website *www.p21.org*; Assessment & Teaching of 21st Century Skills website, *www.atc21s.org*.

As a result, a great deal of the assessment research in recent years has focused on innovative and "authentic" forms of assessment that would be able to capture the type of learning that is valued in today's societies. These alternative forms of assessment are most commonly referred to as performance-based assessment. They may include open-ended tasks such as oral presentations, essays, experiments, projects, presentations, collaborative tasks, real-life cases, problem-solving assignments and portfolios. The main characteristic of performance assessments is that they assess a range of integrated knowledge and skills by asking students to perform a task rather than to provide a correct answer. As such, they are more effective at capturing more complex achievements than closed-ended formats (Looney, 2011b).

Developments in information and communication technologies (ICT) have opened new avenues for the assessment of more complex competencies. Technology-enhanced learning environments may in fact provide tools and systems which recreate learning situations requiring complex thinking, problem-solving and collaboration strategies and thus allow for the assessment of such competencies (European Commission, 2011a). Innovative computer-based assessments may now score student performances on complex cognitive tasks, such as how students go about problem solving, or open-ended performances such as written essays, or student collaboration on constructed response formats (Mislevy et al., 2001). With some assessments, students may receive feedback on their work while they are taking the assessment (Lewis, 1998, in Looney, 2011a).

Issues to be addressed

Despite a high degree of interest in teaching and assessing new forms of learning, information collected in the OECD Review indicates that the use of innovative assessment approaches remains quite limited within the national assessment frameworks of OECD countries. Across the countries reviewed by the OECD, stakeholders reported concerns that assessment practices appeared to lag behind current conceptions of successful teaching and learning. Both national assessments and classroom-based assessments in many countries have remained focussed primarily on reproducing knowledge and applying basic skills, with less attention being paid to measuring complex competencies. Hence, while the curriculum might be competency-based, the assessment system may not adequately capture many of the key objectives of the curriculum. Where this is the case, the assessment system can become a "hidden curriculum" encouraging a narrower approach to teaching and learning (Nusche, forthcoming).

Large-scale central assessments in particular tend to focus on a relatively narrow set of cognitive outcomes. The majority of such standardised assessments are focussed on the areas of literacy and numeracy and rely largely on paper-and-pencil tests done by students individually in a finite period of time. The use of technology in standardised assessments also remains limited across the education systems participating in the OECD Review.[2] While standardised assessments will always be limited to measuring a selected subset of curriculum goals, the assessment of more complex competencies is generally expected to happen in classroom assessment, where teachers can use richer and more in-depth assessment tasks. However, while teacher-based assessment provides opportunities for diverse and innovative assessment approaches, studies from different countries indicate that teachers do not necessarily use such approaches (Crooks, 1988; Black, 1993; Black and Wiliam, 1998; Harlen, 2007).

The limited use of performance-based assessments in large-scale assessments may be explained by concerns about reliability, resources and timescales. There are challenges

related to creating reliable measures of complex competencies, such as problem-solving, creativity and collaboration. Performance-based assessments often tend to have lower comparability of results than standardised paper-and-pencil assessments. Research in some countries has shown that higher-order thinking skills are context and situation specific and that it is difficult to generalise from hands-on performance-based tasks to make judgements about student competencies (Shavelson et al., 1990; Linn et al., 1991). Hence, the use of closed-ended paper-and-pencil tests is often motivated by the need for objectivity, fairness and impartiality in assessment, especially where high stakes are attached. Performance-based assessments are also more costly and time-consuming to implement on a large scale.

While it is generally expected that the assessment of more complex competencies happens on a continuous basis in the classroom, there are in fact a number of challenges for teachers to assess the key competencies outlined in many curricula. First, there is often a lack of clarity on how to translate competency aims into concrete teaching and assessment activities. Competency goals are often stated in a general way with little guidance regarding what exactly teachers are expected to change in their teaching and assessment. Second, the transversal nature of competencies – they tend to involve several subjects or go beyond school subjects altogether – makes it challenging for teachers to see who should be responsible for assessing them and how to fit them within particular subjects or disciplines. Third, the high visibility of standardised assessments may put pressure on teachers to adapt their own assessment to the format used in national tests. Teachers may be tempted to narrow their teaching and assessment in order to best prepare their students for closed-ended national tests, to the detriment of richer more performance-based approaches (Lucas and Claxton, 2009; European Commission, 2011b, Pepper, 2011).

Nonetheless, information collected in the OECD Review also revealed a range of innovative assessment formats that have been introduced – often on a small scale – in many countries and contexts to assess students' progress in acquiring and applying complex competencies. There are a variety of promising approaches to achieving better alignment between competency-based curricula and assessment approaches, both large-scale and classroom-based. These innovative approaches will be explored in more detail throughout this chapter.

Towards a new understanding of assessment

National and international student assessment data points to persistent inequities between student groups from different socio-economic, linguistic and cultural groups within a given country. As classrooms across OECD countries are becoming more and more diverse in terms of student backgrounds and prior learning, teachers are increasingly expected to identify what students already know and can do to in order to respond to the learning needs of individual students. This is to be done on the basis of ongoing assessment activities in the classroom. In this context, the thinking about different assessment purposes has evolved considerably over the past decades. While assessment has traditionally been thought of as separate from the teaching and learning process – for example, a test or examination coming at the end of a study unit –, current policy and practice in many countries emphasises the importance of formative assessment or assessment *for* learning, which should occur as an integrated part of day-to-day classroom interactions (Looney 2011a).

Interest in formative assessment strategies has been fuelled by a great deal of research pointing to the positive impact of such assessment on student learning (Box 4.1). Formative assessment, which emphasises the importance of actively engaging students in their own learning processes, also resonates with countries' goals for the development of students' higher-order thinking skills, metacognition and skills for learning to learn (Box 4.2). It also fits well with countries' emphases on the use of assessment and evaluation data to shape improvements in teaching and learning and is consistent with a focus on creating learner-centred, structured, personalised, social and inclusive learning environments (Looney, 2011a; Istance and Dumont, 2010). In this context, assessment for learning in several countries has become an integral element of the curriculum, and it is understood as an element which can actually enhance, not simply measure, the achievement of the curriculum (Stiggins and Arter, 2007).

The concept of formative assessment is open to a variety of interpretations in assessment policies across different countries. However, despite some contestation around meaning, there is a strong commitment across OECD countries to formative approaches and to developing school practices in this area. As shown in Table 4.A1.2 (Annex 4.A1), the majority of education systems participating in the OECD Review have now developed policy frameworks to support and promote formative assessment in the classroom. Such policy frameworks shift attention from teacher-centred programmes towards the learners themselves, requiring teachers to adapt teaching techniques to meet learning needs and helping students develop their own assessment capacities. While summative assessment and reporting remain essential at key stages of the education process, formative assessment frameworks tend to shift attention away from excessive focus on numerical marks, labelling and ranking of students, in order to focus on learning processes and individual progress.

While formative assessment is mostly about interactions inside the classroom, it is important to note that it is not at odds with external assessment approaches. Quite the contrary, information from external assessments can complement teachers' own assessment strategies and may also be used formatively to identify learning needs and adjust teaching strategies. Several countries have recently developed standardised assessments with a formative purpose, which have no stakes for students. Such assessments, although externally set and often externally marked, are also designed with teachers' assessment practice in mind – they can give teachers an insight into national expectations and standards and provide feedback to students on their progress.

Issues to be addressed

As outlined by Looney (2011a), a long-held ambition for many educators and assessment experts has been to integrate summative and formative assessment more closely in order to build comprehensive and consistent assessment frameworks that balance regular assessment for improvement with punctual assessments for summative and accountability purposes. Currently, however, many countries are facing challenges in combining the new understandings of how students learn with well-established expertise in relation to summative assessment (Irish Department of Education and Skills, 2012).

Evidence from the Country Backgrounds Reports and the country-specific *OECD Reviews of Evaluation and Assessment in Education* indicates that many education systems were struggling to embed a deep understanding of formative approaches in regular classroom practice. In fact, in many settings, formative assessment was understood as "summative assessment done more often" or as practice for a final

summative assessment, rather than being used by teachers and students jointly to reflect on and respond to learning needs. This illustrates a common misunderstanding and misinterpretation of the meaning and intentions behind formative assessment. Formative assessment needs to be independent of the requirement to accredit performance. Its aim should be to identify misunderstandings, misconceptions or missing elements of student learning in order to change instruction and provide detailed feedback.

Also, while giving feedback is a regular aspect of classroom interactions in all countries, not all types of feedback are adequate to promote student learning (Boulet et al., 1990; Butler, 1988; Swaffield, 2008). Across the countries that received individual OECD Country Reviews, many teachers used the following approaches to what they considered formative assessment: drawing students' attention to their mistakes in a test or a task, asking students to make more effort, and giving students praise in order to motivate them. However, research shows that such feedback which does not provide students with specific guidance on how to improve, or that is "ego-involving", even in the form of praise, may in fact have a negative impact on learning (Köller 2001; Mischo and Rheinberg, 1995; Pryor and Torrance, 1998; Swaffield, 2008; Wiliam, 2010).

Many systems are also facing challenges in the effective use of external assessments for formative purposes. In many cases, the data gathered in large-scale assessments are not at the level of detail needed to diagnose individual student needs (McGehee and Griffith, 2001; Rupp and Lesaux, 2006) nor are they delivered in a timely enough manner to have an impact on the learning of students tested. Also, in several countries, tensions have arisen when an assessment is being used for both formative and summative purposes. As explained by Linn (2000), assessment systems that are useful for formative and monitoring purposes usually lose much of their credibility when high stakes for students, teachers or schools are attached to them, because the unintended negative effects of the high stakes are likely to prevail over the intended positive effects (Box 4.1).

Nonetheless, while challenges remain, evidence from the OECD Review points to a number of promising approaches used by schools, regions or entire education systems in order to promote formative assessment and integrate both formative and summative approaches within coherent frameworks for student assessment. Key elements in developing such balanced assessment strategies will be explored throughout this chapter.

Governance

This section deals with the governance of student assessment systems across OECD countries. It describes the different purposes and objectives of student assessment systems and the legal frameworks in place to ensure that student assessment results are used in a way that such objectives are reached. The section also explores how responsibilities for assessment are distributed in different countries and how different levels of governance interact to form a comprehensive assessment system.

Purposes

This sub-section describes how countries define and regulate the different purposes of assessment in their education systems. It explores the policy frameworks for *(i)* summative assessment and *(ii)* formative assessment. While the same processes and assessment formats may be used for both summative and formative purposes, the two approaches differ in the way the assessment results are acted upon. While summative

assessment aims to provide a summary statement about past learning, formative assessment is intended to inform future teaching and learning.

Assessment for summative purposes

As explained above, summative assessment, or "assessment *of* learning" involves judging student performance for a decision or record (Ewell, 2005). It usually occurs at the end of a learning unit, term, school year or educational level (Eurydice, 2009b). The results of summative assessment can be reported in different forms including marks, transcripts, certificates and qualifications. The intentions for designing and implementing summative assessment strategies include:

- **To motivate students** to increase effort and achievement. The marks, transcripts or diplomas connected to summative assessment are often conceived as rewards for having performed successfully on an assessment.
- **To provide information** about student performance to a range of different stakeholders, such as the students themselves, their parents, others within the school, or school-external players such as employers.
- **To select or group students** according to their achievement levels. In many countries, assessment results are used to stream students according to their ability levels within schools, or to select them into certain types of schools.
- **To certify learning and award qualifications** that grant students access to higher education institutions or certain professions.

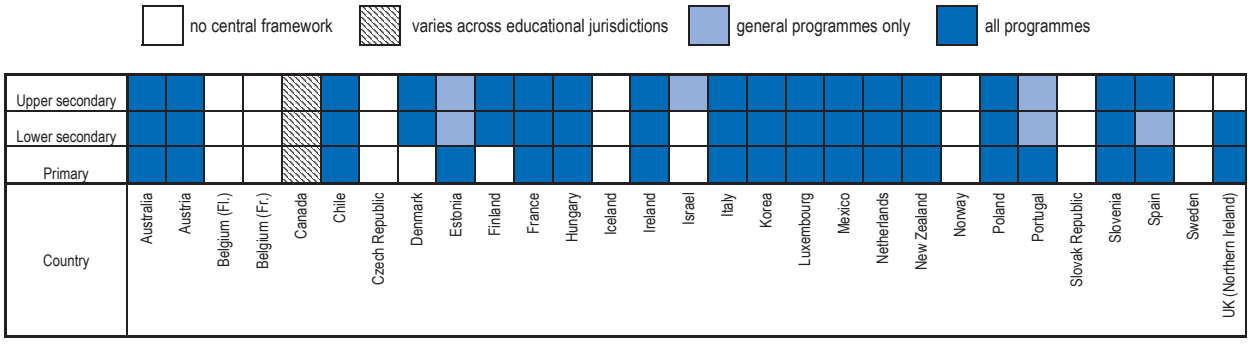
Across the OECD, countries draw on a range of different approaches to implement summative assessment. In the first years of education, summative assessment tends to be the responsibility of school-level professionals in most countries. At this level, summative assessment typically serves for school-internal purposes such as keeping records and giving reports of progress to students, parents and other teachers. Summative assessment for school-external purposes such as selection and certification tends to become more important as students progress to the higher levels of school education (Nusche, forthcoming).

Internal summative assessment

Internal summative assessment, implemented by teachers in the classroom, plays a key role across OECD countries. The majority of education systems have developed policy frameworks (national or state laws or regulations) that specify procedures for internal summative assessment, particularly in secondary education where summative assessment typically receives increased attention (Figure 4.2). These frameworks are generally developed at the central level[3] and they apply to all schools in the majority of education systems.[4]

While policy frameworks for internal summative assessment are commonplace in most countries, there are large variations regarding their level of detail and prescription. In Chile, for example, the policy framework sets very basic requirements, such as the number of assessments per year and requires schools to establish an internal protocol for assessment. In Finland, the requirements for internal summative assessment are included in the national core curriculum for general programmes and in the national qualification framework for vocational programmes. In Ireland, at ISCED levels 2 and 3, subject-specific assessment requirements are provided within the subject syllabi. In Poland, the framework is very general and leaves much autonomy for schools to set up their own assessment rules.

Figure 4.2 Existence of central frameworks for internal summative assessment at ISCED levels 1-3 (2012)



Source: Derived from information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability across countries.

A number of education systems do not have formal frameworks for internal summative assessment at any level of education, but certain basic requirements for summative assessment are typically set in the legislation, curriculum or regulations. For example, in the Flemish Community of Belgium, each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of central attainment targets and developmental objectives, and the Inspectorate may ask schools to present their policy. In the French Community of Belgium, the policy framework is provided through the work of the Inspectorate and the possibilities for students to appeal summative assessment decisions made by the class council. In the Czech Republic, schools are legally required to set their own assessment systems, which are approved by the school board and controlled by the Inspectorate. In Iceland, there is a requirement for students to undergo summative assessment at the end of Year 10, but the curriculum and regulations are flexible regarding how this is implemented by schools and it may in fact take the form of a formative assessment. In Norway, the Education Act states that students shall obtain summative achievement marks at the end of each year level in secondary education.

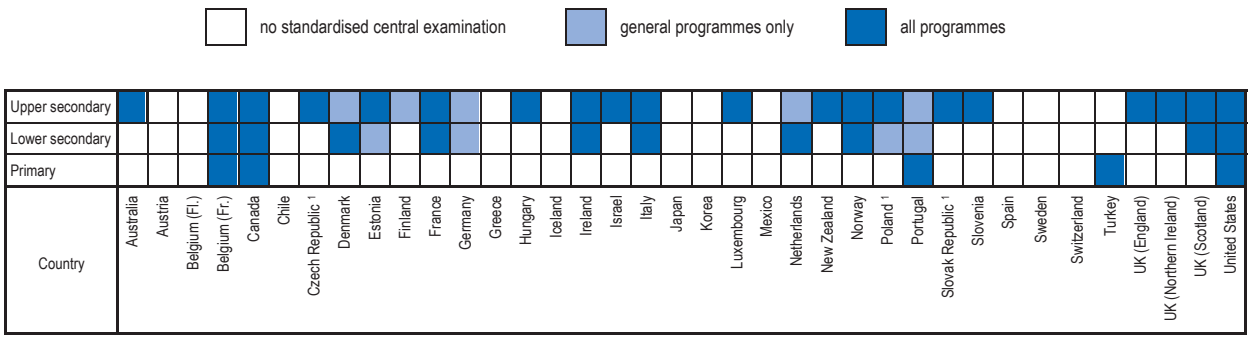
Regarding the primary education sector, it is important to note that some countries, such as Denmark and Norway, have an explicit policy to avoid summative marking and reporting in the first years of education, which explains the absence of a summative assessment framework at this level (Nusche et al., 2011a; Shewbridge et al., 2011).

Central examinations

In addition to internal summative assessments, many OECD education systems use central examinations, i.e. standardised assessments that are developed at the central level and have a formal consequence for students (e.g. influence on a student's eligibility to progress to higher levels of education) to measure student performance. Such examinations are rarely used in the early years of schooling and become more widespread at the higher levels. Of 37 OECD education systems for which information was available, only five used central examinations at the primary level (the French Community of Belgium, Canada, Portugal, Turkey and the United States), versus 14 education systems at the lower secondary level and 25 at the upper secondary level (Figure 4.3) (OECD, 2012a, complemented with information collected from countries participating in the OECD Review). In addition, standardised examinations offered to schools by private providers play an important role in some countries. In the Netherlands, for example,

85% of primary schools use the school leavers test developed by the Central Institute for Test Development (Cito), which provides information regarding the school type most suitable for each student in the next phase of education.

Figure 4.3 Existence of standardised central examinations at ISCED levels 1-3 (2012)



Notes: (1) Excludes ISCED 3C programmes at the upper secondary level.

Source: OECD (2012a), *Education at a Glance 2012*, complemented with information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

Assessment for formative purposes

As illustrated above, formative assessment, or "assessment *for* learning", aims to deepen and shape subsequent learning rather than making a judgement about past performance (Black and Wiliam, 1998). It is essentially a pedagogical approach consisting of frequent, interactive checks of student understanding to identify learning needs, provide feedback to students and adapt teaching strategies (OECD, 2005a). It is embedded in the normal day-to-day teaching and learning process and may include activities such as classroom interactions, questioning and feedback (Looney, 2011a). The use of assessment information is key to the concept of formative assessment: to be considered formative, assessment evidence must be acted upon in subsequent classroom teaching and learning. Students also participate actively in the process through self- and peer-assessment. In some recent formulations, the active participation of students in the process has given rise to the term *assessment as learning*, which focuses on students reflecting on and monitoring their own progress to inform future learning (Earl, 2003). The intentions for designing and implementing formative assessment strategies include:

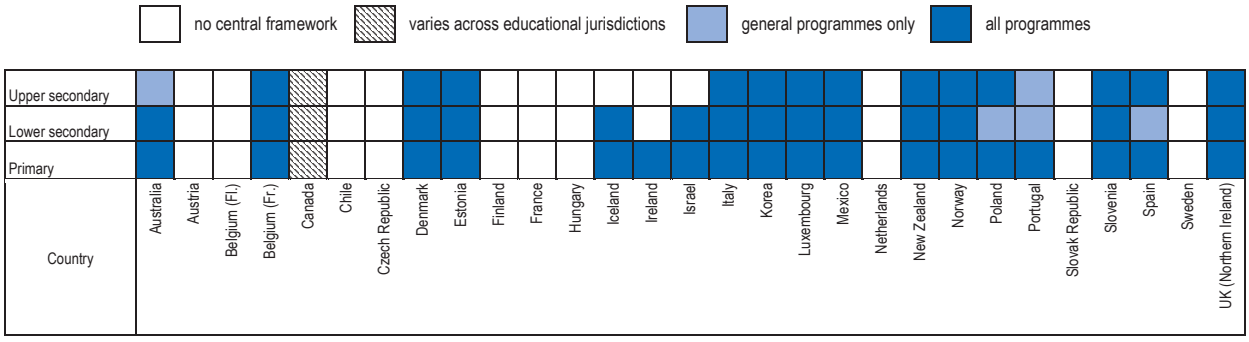
- **To provide timely feedback to students**, which they can integrate into their learning process. Several studies indicate that feedback is most effective when it is timely, is tied to criteria regarding expectations, and includes specific suggestions for how to improve future performance and meet learning goals (Wiliam, 2006; Hattie and Timperley, 2007; Swaffield, 2008).
- **To help students to feel safe to take risks and make mistakes** in the classroom. Students are thus more likely to reveal what they do and do not understand and are able to learn more effectively (Looney, 2011a).
- **To diagnose student learning needs and differentiate teaching accordingly**. In order to develop an appropriate teaching intervention, teachers need to assess students' learning needs and explore a range of potential causes of learning difficulties (Looney, 2011a).

- **To actively engage students in their own learning processes** so as to develop higher-order thinking skills and skills for "learning to learn", and to allow students and teachers to engage in conscious reflection on the learning process (Earl, 2003).

Internal formative assessment

Given the widely reported benefits of formative assessment for the improvement of teaching and learning, many OECD education systems have developed policy frameworks (national or state laws or regulations) to promote and support formative assessment practice in the classroom (Figure 4.4). Where these frameworks exist, they tend to be developed at the central (national or state) level and apply to all schools.[5] The existing frameworks generally include a requirement for schools to implement formative assessment in the classroom. In Australia (ISCED 2 and 3 only), Korea and Mexico, they also include a requirement for formative assessment to be part of initial teacher education programmes. In Korea, there is also a requirement for teachers to undertake professional development in this area. In Estonia, it is mandatory for schools to report on their strategies to promote formative assessment. In Spain, the regulations are most extensive including a requirement for schools to implement student formative assessment and to report on their strategies to promote student formative assessment, as well as for student formative assessment to be part of initial teacher education programmes and for teachers to undertake professional development in this area.

Figure 4.4 Existence of central frameworks for formative assessment at ISCED levels 1-3 (2012)



Source: Derived from information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

In some education systems, while formative assessment is not inscribed in national or state education law, it is promoted through other documents. In the Flemish Community of Belgium, primary schools are required to monitor the progress of every student and report observations to parents, but there are no specific regulations regarding the procedures for doing so. In Hungary, elements of formative assessment such as verbal assessment and differentiated assessment methods are included in legal regulations and the national core curriculum. In Ireland, the National Council for Curriculum and Assessment (NCCA) has issued guidelines to secondary schools that emphasise the value and use of formative assessment and while they are not in the form of regulations, they will be a key part of assessment requirements for the new ISCED 2 curriculum to be introduced on a phased basis from 2014. In Finland, the national core curricula for all ISCED levels mention that teachers should observe students' progress. In the Netherlands, draft laws are currently being prepared to set a requirement for schools to use formative assessment systems for

results-based work in schools. A number of countries, including Austria, Chile, the Czech Republic, France, Hungary, the Slovak Republic and Sweden do not have specific central regulations or documents promoting formative assessment.

It should also be noted that while existing policy frameworks signal the high level of attention given to formative assessment at the policy level, little information is available regarding the effective and systematic implementation of formative assessment across schools and classrooms. In a number of education systems participating in the OECD Review, the understanding and development of formative assessment appears to be still at an early stage of development. To ensure that policy commitments to formative assessment are matched with actual developments in the classroom, sustained investment in teachers' understanding and capacities regarding formative assessment is necessary. Box 4.3 provides some examples.

Box 4.3 Matching a commitment to formative assessment with concrete support for teachers

In **Canada**, many school districts offer professional development opportunities for teachers to improve their skills and knowledge of assessment/evaluation mechanisms. For example, over the past two years in particular, there has been a strong emphasis on Assessment for Learning practices in Nova Scotia schools. To that end, there was a provincial assessment summit in 2009 and several Boards then hosted their own Assessment Summits in 2010. The South Shore Regional School Board in Nova Scotia hosted a two day event in September 2010. As well, Assessment for Learning has been a Board priority in its Educational Business Plan and it remains so today. A website on assessment has been designed for teachers providing a multi-media workshop on the full scope of assessment knowledge, skills and applications (*http://web.ssrsb.ca/assessment/*).

In **Ireland**, the National Council for Curriculum and Assessment (NCCA) has contributed to the development of expertise in formative assessment through its curriculum development projects with schools. As part of its work with groups of teachers in its Primary School Network, the NCCA explores how formative assessment approaches can be implemented in Irish classrooms. The NCCA has also designed materials to support teachers and schools in expanding their assessment toolkit. Its Assessment for Learning website includes multi-media support and materials such as classroom video footage and samples of children's work with teacher commentary. There are also reflection tools and checklists to support individual teachers and whole school staffs in reviewing current assessment practice (*http://action.ncca.ie/primary.aspx*).

In **Norway**, a statutory requirement has been introduced for schools to implement assessment for learning. To support teachers in fulfilling the requirements for formative assessment, the Directorate for Education and Training has created a website on assessment for learning providing a range of materials and tools including questions for reflection, films, assessment tools and literature, and also examples of different ways to document formative assessment practice. At the same time, there has been a developing awareness that teachers have not traditionally received training in formative assessment and that there was very little expertise available nationally for school leaders to draw on to provide support. To address this, the Ministry of Education and Research and the Directorate for Education and Training in Norway identified formative assessment as a priority area for education policy and professional development and launched a range of support programmes and learning networks at the regional, local and school level. For example, the Assessment for Learning programme (2010-14) is organised in learning networks at the local and regional level, where practitioners can exchange experience and create spaces for common reflection on effective practice. Participating municipalities and counties employ a formative assessment contact person to assist in running the project locally. These contact persons attend Assessment for Learning workshops run by the Directorate. The programme also provides online resources including tools and videos on how to enact effective formative assessment in the classroom.

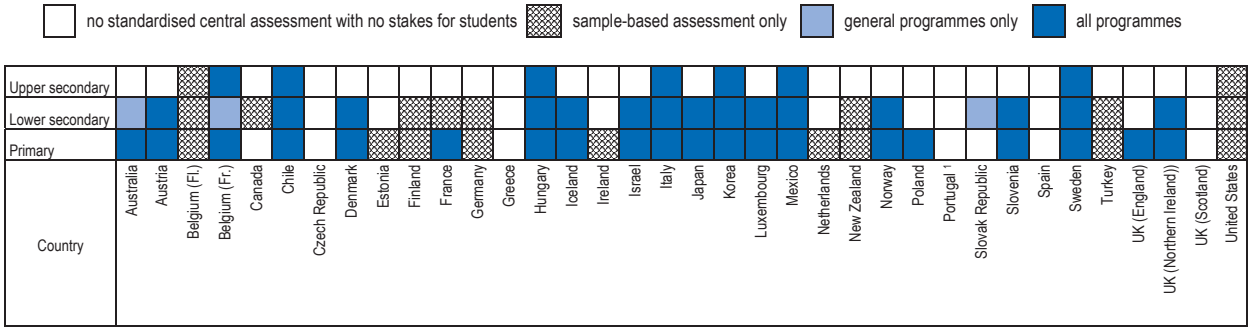
Sources: Fournier and Mildon (forthcoming); Irish Department of Education and Skills (2012); Nusche et al. (2011a).

Central assessments for formative use

Information from standardised assessments may also be used formatively to identify learning needs and adjust teaching strategies. Several education systems have developed standardised central assessments that do not have stakes for students. In some cases, these assessments were developed with the explicit purpose of providing formative or diagnostic information to teachers. In other cases, they are primarily aimed at producing aggregated performance information to monitor education system performance. The features of central assessments for system monitoring will be discussed primarily in Chapter 8 on system evaluation, but references will be provided in this chapter, because the results of such assessments may also be fed back to schools and be used by teachers in a formative way. Even where the central assessments test only a sample of students, the results may be provided to the participating schools.

Figure 4.5 below provides an overview of central assessments with no stakes for students across education systems. Clearly, such central assessments are most common at the primary (29 education systems) and lower secondary level (27 education systems), and less common at the upper secondary level (9 education systems). In most education systems, they assess the full cohort of students in the relevant years. In a number of education systems, these central assessments are sample-based only (Flemish Community of Belgium, Canada, Estonia, Finland, Germany, Ireland, the Netherlands, New Zealand, Turkey and the United States). France also has sample-based assessments at ISCED levels 1 and 2, in addition to full-cohort central assessments at ISCED level 1. England in the United Kingdom also has a sample-based assessment of science at ISCED level 1, in addition to full-cohort assessments of English and mathematics. In Ireland, all schools at ISCED level 1 are required to implement annual standardised assessments developed by private providers and a sample of primary schools participates in national assessments on a 5-year cycle.

Figure 4.5 Existence of standardised central assessments with no stakes for students at ISCED levels 1-3 (2012)



Note: (1) Before 2012/13 there were national assessments in Portuguese and mathematics.

Source: OECD (2012a), *Education at a Glance 2012*, complemented with information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

Using large-scale assessments for formative purposes

Low stakes central assessments provide external signposts for teachers and students by indicating the learning goals that are expected nationally and can offer interesting pedagogical tools for teachers. In several countries where OECD Country Reviews took place, teachers were positive about such formative assessments and saw them as a tool to help them decide what they should be focusing on in their improvement plans to support

individual students. Positive effects of using student results from large-scale assessments to inform teaching may include: greater differentiation of instruction, greater collaboration among colleagues, an increased sense of efficacy and improved identification of students' learning needs (van Barneveld, 2008).

At the same time, in many settings there appear to be barriers to the effective use of large-scale assessments in the classroom. An important concern is that teachers often receive the results from such assessments too late to use them pedagogically. In fact, assessment data appear to have the most impact on student achievement when delivered in timely manner (Wiliam et al., 2004). Data from large-scale assessments, however, are sometimes available to teachers several weeks to months following the actual test day.

Moreover, data gathered in large-scale assessments are often not at the level of detail needed to diagnose individual student needs. Large-scale assessments typically include too few items on each topic for teachers to understand where students need help and what strategies might be useful in supporting them (Rothman, 2010). Especially for assessments that are also used for national monitoring, the scores are typically tied to broad proficiency categories, such as: below basic, basic, proficient, advanced (McGehee and Griffith, 2001). It has been argued that these categories are too broad to provide any kind of diagnostic information necessary for profiling individual student needs. In a study on the relationship between performance on a standards-based reading assessment and performance on a battery of diagnostic assessments, Rupp and Lesaux (2006) found that the standards-based assessment provided only weak diagnostic information, and masked significant heterogeneity in the causes of poor performance.

While large-scale standardised assessments can be useful to provide some initial clues about areas that need attention, other additional diagnostic assessments are needed to identify the causes of poor performance and develop an appropriate instructional intervention. Overall, while there is some evidence that data from large-scale assessments are being used successfully to identify students' strengths and weaknesses, to change regular classroom practice or to make decisions about resource allocation (Anderson et al., 2004; Shepard and Cutts-Dougherty, 1991), they need to be embedded in broader, more comprehensive assessment systems that include a range of summative and formative assessments, curriculum tasks, instructional tools, and professional development that helps teachers understand which assessment information is most appropriate for a particular purpose.

Balancing formative and summative assessment purposes

Finding a balance between formative and summative assessment is a challenge shared by many education systems. It is made more complex by the wide range of understandings of the meaning of the term "formative assessment", and the difficulty of managing the tensions between a stated commitment to formative assessment on the one hand, and public, parental and political pressure for accountability in the form of scores and rankings on the other (Harlen and James, 1997; Newton, 2007). This pressure for summative scores, and a conflation of formative and summative purposes in education policy documents sometimes results in confusion that in some cases may have hindered sound assessment practice, especially in the development of formative assessment (Harlen and James, 1997; Newton 2007). While the attention to results and data is a positive feature of education systems, an over-emphasis on these may have a negative impact, and undermine the formative role of teachers and assessment often so highly valued in policy goals.

As discussed above, there are cases in which there are misunderstandings and misinterpretations by teachers of the meaning and intentions behind formative assessment. In many education systems, teachers have long held the main responsibility for classroom assessment and see assessment as an important professional responsibility. At the same time, it is often not well understood that assessment for learning requires a major shift in mindset for teachers, as well as fundamental changes *vis-à-vis* traditional classroom assessment practices. While continuous classroom assessment, done by teachers on a regular basis, can include both summative and formative assessment, these labels represent fundamentally different purposes. Formative assessment is the process of identifying aspects of learning as it is developing, so that learning itself can be enhanced. It needs to be separate from the process of rating and ranking performance.

The challenge is to ensure that teachers move beyond surface techniques for formative assessment (such as "summative assessment done more often" or feedback that is unspecific or ego-involving) in order to adopt effective assessment for learning approaches. To have the greatest impact, feedback needs to provide information not only on how the learner has done, but also on the specific steps needed to progress further. It needs to be timely, detailed and specific (Wiliam, 2006; Hattie and Timperley, 2007; Swaffield, 2008). This requires ongoing professional learning opportunities and support for teachers, as well as suitable strategies to integrate classroom-based formative assessment within a broader framework for student assessment.

Developing coherent assessment frameworks, where each assessment is fit for purpose

In many education systems participating in the OECD Review, the introduction of central examinations and assessments is relatively recent and in some instances the communication about their purposes has not been sufficiently clear. If a specific assessment was designed for a particular purpose, its use for a different purpose is not ideal because it is likely that inferences made based on the test result will not be accurate, valid or useful for other purposes. For example, if the purpose of an assessment is diagnostic, then the test needs to be designed to provide very fine-grained information which allows uncovering specific difficulties and misconceptions of individual students. On the other hand, if the purpose of assessment is to compare schools or regions, the test needs to be designed to provide highly reliable summative scores in broad comparable categories (such as the number of students meeting or not meeting standards). It also needs to be complemented by broader information about the school's context, characteristics and processes (see Chapter 6).

There are risks in using a single test for too many purposes, in particular where the information ideally required in each case is not the same (Morris, 2011). In some countries participating in the OECD Review, national assessments have been introduced as low stakes and formative assessments without impact on individual students, teachers or schools. Subsequently, however, additional uses were added to the test, for example they were used to appraise teachers or to monitor school performance. Such shifting purposes may undermine the credibility of the assessment and jeopardise the constructive use of the test's results.

Since assessments need to be designed in line with their specific purpose, it is important for policy makers and assessment developers to be clear from the beginning about the main purpose for which assessment results will be used. According to Newton (2007), where multiple purposes are intended to be achieved, it is essential to clarify the *primary purpose* of

the assessment and to prioritise explicitly among purposes. The primary purpose and, to the extent possible, other high priority purposes should then determine assessment design characteristics to ensure the validity of results for intended purposes.

Newton (2007) suggests that the use of results for more than one purpose is possible if the purposes are not logically incompatible and if there is clarity about the type of evidence the assessment can provide. For example, an assessment that is designed primarily to help education policy makers monitor the education system should also be of value to those who participate in the assessment, and the results should be fed back to them along with illustrations of the types of inferences they can and cannot draw from the results. However, an assessment that is designed primarily for diagnostic and formative purposes should not be used for summative or accountability purposes because this may undermine its primary function (Linn, 2000).

Box 4.4 Defining and communicating the purposes of assessment

In **Canada**, the *Principles for Fair Student Assessment Practices for Education in Canada* outline key elements for assessment practice that have served as foundations for teacher handbooks, board polices and departments of education policy documents on assessment and test development in all Canadian jurisdictions. The Principles were developed in response to what was perceived as assessment practices not deemed appropriate for Canadians students. These principles and guidelines intended for both assessment practitioners and policy makers identify the issues to be taken into account in order that assessment exercises to be deemed fair and equitable. The text acts both as a set of parameters and a handbook for assessment. The first part deals with developing and choosing methods for assessment, collecting assessment information, judging and scoring student performance, summarising and interpreting results, and reporting assessment findings. It is directed towards practising teachers and the application of assessment modes in the classroom setting. The second part is aimed at developers of external assessments such as jurisdictional ministry/department personnel, school boards/districts, and commercial test developers. It includes sections on developing and selecting methods for assessment, collecting and interpreting assessment information, informing students being assessed, and implementing mandated assessment programs (for more information, see: *www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf*).

In **Denmark**, official information on the national tests produced by the former School Agency clearly repeats the message that the national tests only measure a discrete area of student knowledge and skills and teachers should use a range of other tests to gauge student progress. For example, it is stressed that the Danish test only measures students' proficiency in reading and a wide range of key knowledge and skills in Danish (e.g. spelling, grammar, punctuation, cultural understanding, literary knowledge, ability to express oneself) is not tested. Educators are aware that the tests provide only a snapshot of students' achievement levels in select learning targets and subjects (Wandall, 2010).

The **New Zealand** Ministry of Education *Position Paper on Assessment* (2010) provides a formal statement of its vision for assessment. It describes what the assessment landscape should look like if assessment is to be used effectively to promote system-wide improvement within, and across, all layers of the schooling system. The paper places assessment firmly at the heart of effective teaching and learning. The key principles highlighted and explained in the paper are: the student is at the centre; the curriculum underpins assessment; building assessment capability is crucial to achieving improvement; an assessment capable system is an accountable system; a range of evidence drawn from multiple sources potentially enables a more accurate response; effective assessment is reliant on quality interactions and relationships. To support effective assessment practice at the school level, the Ministry of Education is also currently conducting an exercise which maps existing student assessment tools. The purpose is to align some of the assessment tools to the National Standards and provide an Assessment Resource Map to help school professionals select the appropriate assessment tool to fit their purpose.

Sources: Fournier and Mildon (forthcoming); Shewbridge et al. (2011); Nusche et al. (2012).

In this context, it is important that all stakeholders have a broad understanding of assessment and of the need to combine a range of different assessment information in order to come to good judgements and valid interpretations on student learning and progress (Absolum et al., 2009). Teachers need skills to interpret standardised assessment results, to understand where further diagnostic testing of some students is necessary and to identify areas were teaching strategies may need adjustment to meet student needs.

A key governance challenge for countries is to develop a clear vision and strategy for assessment where different approaches developed nationally and locally each serve a clearly defined purposes and the format of the assessment is aligned to these particular purposes. In New Zealand, for example, the Ministry of Education has published a position paper on assessment in 2010, outlining the underlying principles of assessment. This has been complemented by a mapping exercise to clarify the distinct aims of existing standardised assessment tools that are at teachers' disposal (Box 4.4). Also, clear communication is vital to ensure that assessment results are used in an effective way. In particular, it is important to establish safeguards to avoid an over-emphasis on standardised assessment results. One such safeguard is to communicate clearly about the kinds of evidence that different types of assessment can and cannot provide (for examples from Canada and Denmark, see Box 4.4).

Responsibilities for student assessment

This sub-section provides an overview of the different actors involved in student assessment and the ways in which the different levels of governance interact. Responsibilities for student assessment are typically shared between different agencies and levels of the education system. In most countries, teachers have the main responsibility for continuous formative and summative assessment in the classroom, whereas regional, state or national agencies tend to be in charge of developing and marking standardised assessments that are used at key stages of education.

Internal assessment

In most OECD countries, schools benefit from considerable autonomy in the organisation of internal student assessments. School leaders, together with teachers, and sometimes in co-operation with school governing boards or education authorities, are typically in charge of establishing school policies for student assessment. Across the OECD, in PISA 2009, 66% of 15-year-old students were in schools whose principals reported that the school alone had the main responsibility for establishing student assessment policies, and 23% of students were in schools where the school together with the regional and/or national education authority had considerable responsibility for student assessment policies (OECD, 2010a).

While schools tend to have considerable freedom in establishing their own assessment policies, certain basic requirements are generally set in central policy frameworks (see above). The frameworks for internal summative assessment are centrally defined in most education systems, but in fact, different levels of education are involved in ensuring compliance with these frameworks. Across OECD education systems, this task may be attributed to the school leaders and teachers (e.g. Chile, Estonia, Finland, France, Hungary[6], Israel, Poland, Slovak Republic, Slovenia), the subject committees (Korea), the schools boards (e.g. Czech Republic, Iceland, Ireland and New Zealand), the school organising bodies (Hungary, Netherlands), the local education authorities (Norway), the central, state or provincial authorities (Australia, Austria, Canada, Luxembourg, Mexico,

New Zealand) or the Inspectorates (Flemish and French Communities of Belgium, Czech Republic, Netherlands) (Tables 4.A1.1, 4.A2.1a, 4.A2.1b, Annexes 4.A1 and 4.A2).

In their national background reports prepared for this Review, the majority of education systems reported about long-standing traditions of teacher-developed assessment and the historically important role of teachers' professional judgements in assessment. Teachers are generally expected to take responsibility for different functions of assessment including diagnostic, formative and summative. While teachers tend to have the exclusive responsibility for summative assessment in primary education, their assessment approaches are typically complemented by regionally or nationally implemented standardised examinations at the higher levels of education (Figure 4.3). The distribution of responsibilities tends to be organised in a way that teachers assess and report on student performance in relation to the full range of curriculum goals, while standardised examinations and assessments assess a particular subset of learning goals in specific year levels.

Regional and central examinations

Tables 4.A1.4, 4.A2.4a and 4.A2.4b (Annexes 4.A1 and 4.A2) provide detailed information about the groups involved in national student examinations that have a formal consequence for individual students. Student examinations are considered "standardised" if they are designed to ensure that the questions, conditions for administering, marking procedures, and interpretations are consistent and comparable among students (Popham, 1991). To ensure these conditions, examinations are often designed and marked at a central level. At the same time, many countries have hybrid forms where assessments are centrally developed but locally administrated and/or marked. In this case, countries tend to use guidance materials and moderation to ensure the reliability of local marking.

At the lower secondary level, the central education authorities have full responsibility for developing national examinations in the French Community of Belgium, Denmark, France, Italy and Norway, while they share responsibility for this task with a central agency for assessment in Australia, Estonia and Mexico. In some countries, including Ireland, the Netherlands and Poland, central/regional agencies for assessment hold the main responsibility for the development of external examinations. While the examinations are centrally developed in all education systems, school level examiners play a key role in the marking process in several education systems including the French Community of Belgium, Denmark, Estonia, France, Italy and the Netherlands. Where high stakes examinations are marked locally by teachers, moderation arrangements are typically in place to ensure the reliability of marking, for example through the involvement of a second marker in addition to the students' own teachers (more on this below).

At the upper secondary level, the central education authorities have full responsibility for developing national examinations in the French Community of Belgium, Denmark (general programmes only), France, Hungary (general programmes only), Israel, Italy, Luxembourg, New Zealand and Norway. They share responsibility for this task with a central agency or institute for assessment in Australia, the Czech Republic, Estonia and Mexico. Central or regional agencies for assessment hold responsibility for developing examinations in general education programmes in Finland, Ireland, the Netherlands and Slovenia, and for developing examinations in both general and vocational programmes in the Slovak Republic and Poland. In vocational programmes, the schools themselves play a key part in developing examinations in Denmark and Slovenia.

The examinations in upper secondary education are centrally marked by the education authorities and/or a central agency in most education systems. At the same time, school level examiners play a key role in the marking process in a range of education systems including the French Community of Belgium, the Czech Republic, Denmark, France, Italy, Luxembourg, the Netherlands and Slovenia (vocational programmes only). In Hungary, the marking is undertaken at the school level for the "normal"-level examination and at the central level for the advanced-level examination. In New Zealand, internally assessed subjects are marked in the school whereas externally assessed papers are marked by a central agency (the New Zealand Qualifications Authority). In Norway, centrally-given written examinations are externally marked whereas locally-given oral examinations are marked together by the subject teacher and an external examiner. Examinations in vocational programmes are marked by the local education authorities.

Assessments developed by other providers

Private providers play an important role in test development in several education systems. No internationally comparable information is available regarding the importance of private testing companies across countries. From the OECD Country Reviews, it appears that the use of private tests by schools is commonplace in most countries. In some education systems, such as Ireland (ISCED 1 at present and ISCED 2 from 2014), schools are required to choose from among certain standardised tests developed by private providers and to report the results to their school boards and the educational administration.

In the Netherlands, schools are required to report on students' learning results in the final phase of primary education in a way that clarifies the extent to which students have reached the minimal achievement levels for primary education. While schools are free to use different assessment instruments for this purpose, the vast majority of primary schools use to this end the school leavers test developed by the Central Institute for Test Development (Cito). The results from this test also provide information on the school type which is most suitable for each student in the next phase of education.

In Northern Ireland in the United Kingdom, upper secondary schools are expected to choose standardised central examinations from a range of Awarding Organisations. The relevant Awarding Organisations are responsible for marking the external assessment and where internal assessment is used, they also moderate teachers' marking.

In other countries, such as the Czech Republic, private companies offer testing services that schools can choose to buy to support their regular assessment practice. In several education systems, such as the Flemish Community of Belgium and the Slovak Republic, classroom-based assessment is heavily influenced by textbooks, many of which contain tests.

The potential influence of private assessment companies on teacher assessment practices is substantial. Commercial tests available to schools typically aim to provide summative data of the students' level of knowledge in different subjects. In some countries, private providers also offer tests explicitly designed for formative and didactic purposes (for an example from the Netherlands, see Box 4.5). Schools receive feedback of results providing information on areas for individual students to improve. Certain commercial tests may enjoy direct support by local education authorities and consequently may allow the feedback to schools of comparative data, as schools throughout the municipality use these tests. In the Flemish Community of Belgium, school umbrella organisations make tests available to schools belonging to their network. As such, these tests provide information on each student's competencies, but also a

benchmark for schools to compare to other schools within the network. The use of additional externally developed tests can provide schools with useful information as part of their wider student assessment systems.

Box 4.5 Monitoring student learning in the Netherlands

Since the mid-eighties primary schools started to make use of a pupil monitoring system, the LVS (*Leerling Volg Systeem)* developed by the Central Institute for Test Development (Cito). Later on pupil monitoring systems were also implemented in secondary schools and currently every secondary school has a pupil monitoring system. The Cito pupil monitoring system (LVS) for primary education is a consistent set of nationally standardised tests for longitudinal assessment of a pupil's achievement throughout primary education, as well as a system for manual or automated registration of pupil progress. The LVS covers language, (including decoding and reading comprehension), arithmetic, world orientation (geography, history, biology), social-emotional development, English, science and technology. It is purchased by schools at their own cost and initiative. The primary objective of the LVS is the formative assessment of student achievement and individual students' mastery of key subject matter areas in relation to their year level. Item Response Theory is used to vertically equate students' scores in the LVS tests, which allow for a calculation of student growth trajectories in primary school. Since 2003, the LVS also contains computer-based tests, some of which are adaptive. The following presentation formats are made available on the basis of the LVS:

- The *pupil report,* which is a graph in which the pupil's progress is visible throughout the years. Data available in the national surveys are used as a frame of reference, based on percentiles, so that the position of an individual pupil with regards to five reference groups (25% highest scoring pupils, just above average, just below average, far below average, and the 10% lowest scoring pupils) is immediately visible from the corresponding graph.
- For children with special education needs, and who visit special education schools, an *alternative pupil report* is made available. This report also shows at what level a pupil is functioning and how to interpret the results of the pupil compared to children of the same age who attend mainstream primary education.
- In the so called *group survey* the results of all the pupils from a group over a number of years are presented in a table. For each pupil the scale of ability score at the successive measuring moments is shown along with the level score.

Source: Scheerens et al. (2012).

However, in some countries participating in the OECD Review, there are a number of concerns related to the regular use of private tests in the classroom. Often, the most widely used commercial tests contain mainly multiple-choice items and closed-format short answer questions, which are best suited to assess knowledge-based elements of the curriculum. In some cases, commercial tests were being perceived as practice tests for national assessments, hence reinforcing a focus on assessing only a limited subset of learning objectives that can be computer-scored. It was not always clear to the OECD review teams that the tests offered to schools were closely aligned with national curricula. Also, teachers who use commercial tests irrespective of the national and local education goals and without eliminating non-relevant content may present their students with too much content at the expense of essential national or local learning objectives. It is important that independent information about the quality and relevance of private tests is made available to teachers including to what extent they offer useful feedback on student progress against the national learning objectives. Ideally, there should be an accreditation process to validate the use of such tests as reflecting national student learning objectives.

Balancing external assessments and teacher-based assessments

Many countries rely on a mix of external and internal assessment, but finding the right balance between the two approaches may be challenging.

The major advantage of external standardised assessment is its high reliability. It ensures that all students are assessed on the same tasks and that their results are measured by the same standards. Standardised external assessment is usually conducted in supervised conditions which ensure that what is assessed is the students' own work (Crooks, 2004). It is marked by a machine or by external assessors and the marking criteria are standardised, so that a high degree of reliability is given. The marking is expected to be free of bias or discrimination, as the assessors do not know the students whose work they are reviewing. The results are made as objective as possible so that they are, within a year, comparable among students, regardless where they go to school (Rosenkvist, 2010). Externally defined assessments can clearly indicate the standards that are expected nationally of all students, so that they can steer their learning in that direction (Elshout-Mohr et al., 2002).

However, external assessment is often criticised for having lower validity than teacher-based assessment. It tends to be in the form of a written test under supervised conditions, so that only a limited range of curriculum goals can be covered. Also, external assessment typically takes place on very few occasions and thus gives limited information about students' competencies due to the normal daily variations in performance. It can also have detrimental effects on teaching and learning, as teachers may end up focussing on test-taking skills, especially when high stakes are attached to the test results (Box 4.1). The high stakes that are often attached to a single external examination can cause stress or test-anxiety among students resulting in their achievements being reduced on the examination day (Crooks, 2004).

Internal assessment also has its advantages and drawbacks. Due to its continuous nature, teacher-based assessment allows for important achievements to be measured that are more difficult to capture in an external examination, such as extended projects, practical assignments or oral work. Internal assessment thus has a higher potential for the full range of curriculum goals to be covered (Crooks, 2004; Harlen, 2007). As internal assessment is embedded in the regular coursework and spread throughout the course, it is also more authentic than a test-based external examination, providing more opportunities for students to show what they know and are able to do in normal conditions. Especially if particular outcomes are assessed several times, atypical performances as well as achievement trends can be identified (Crooks, 2004).

However, it is important to note that the validity of teacher-based assessment depends to a large extent on the assessment opportunities provided by individual teachers. It is difficult to ensure that all teachers indeed use the potential of internal assessment to cover the full range of goals specified in the curriculum (Harlen, 2007). Several reviews of research on teacher-based assessment note that teacher-made assessments are often no more diverse or innovative than external assessments, encouraging rote learning and recall of fragmented knowledge rather than critical thinking and deeper learning (e.g. Crooks, 1988; Black, 1993; Black and Wiliam, 1998; Harlen, 2007). For internal assessment to work well, it is essential to ensure that teachers receive adequate training to develop their assessment skills (see section on "Capacity").

Internal assessment is often perceived as being less reliable than external assessment. Assessment items and marking standards may vary widely between teachers and schools,

so that the results of internal assessment will lack external confidence and cannot be compared across schools. It cannot always be verified that what is assessed is indeed the student's own work, as some tasks for assessment (e.g. homework and project work) may take place outside the classroom with little supervision (Crooks, 2004). Several studies also report that there is a risk of conscious or unconscious bias in teacher-based assessment, i.e. that teachers may give more help to some students, or in their marking may give inappropriate weight to prior knowledge and expectations of particular students (Crooks, 2004; Harlen, 2007).

Where teachers are responsible for summative assessment, there is also a risk that teachers become subject of parental pressure to lower assessment standards and provide higher marks for their children (Bishop, 2006). Figlio and Lucas (2004) find that parents do not perceive tougher teachers to be better teachers and tend to prefer high marks over high standards. This could lead to potential distortions of results due to parental pressure. With internal summative assessment, the teacher also acquires a double role of teacher and assessor. Rather than strengthening the teacher-student relationship, this may in fact result in a distancing between the student and the teacher (Bishop, 2006). Indeed, students may refrain from asking questions fearing that this could be interpreted as a sign of slow progress and low achievement (Somerset, 1996). The teachers' role in internal summative assessment may thus negatively impact on the effectiveness of their formative assessment approaches.

Research describes several ways to address potential bias in teachers' assessment and increase the reliability of the assessment. There is evidence that the reliability of teacher-based assessments can be improved by the use of scoring guides detailing descriptions of competency levels and providing examples of high performance (Harlen, 2004; 2005). There are also indications that teachers apply assessment criteria more accurately if they are clear about the goals to be achieved and especially if they have participated in the development of criteria (Hargreaves et al., 1996; Frederiksen and White, in EPPI, 2004). External benchmarks showing what is considered to be normal or adequate progress of students in particular marks and subjects are also helpful to help teachers make accurate judgements. In Sweden, for example, teachers are encouraged to compare the achievements of students in internal assessment to student results in national assessments and to use the national assessments as an external guidance and reference points (Nusche et al., 2011b). Finally, training for teachers, teacher collaboration in assessment and external moderation of teacher-based assessment can further enhance the reliability of internal assessments (more on this in the section on "Capacity").

Crooks (2004) suggests that a combination of teacher-based and external assessments would be most suitable to ensure maximum validity and reliability. Learning outcomes that can be readily assessed in external examinations should be covered this way, whereas more complex competencies should be assessed through continuous teacher-based assessment. Where teacher-based assessment is used for summative purposes, it is essential to pay attention to maximising reliability, by using scoring guides, negotiated scoring criteria, external benchmarks, training for teachers, multiple judgements and external moderation. It is also important to provide a range of nationally validated assessment tools that teachers can use to assess their students reliably when they see fit.

Reference points for student assessment

Clear and explicit expectations for student learning and assessment criteria are important to ensure the validity, reliability, transparency and usability of assessment (for

definitions, see section on "Analytical approach"). Information on expected learning outcomes and developments are typically expressed in national curricula, educational standards or learning progressions.

- **National curricula** typically describe overarching learning objectives for the education system and explain the underlying values and culture that should shape teaching and learning. Countries take different approaches to how they design curricula. While some describe the teaching content, methods, materials and assessment criteria to be applied in different subjects and year levels, others establish broad guidelines, leaving room for local authorities and schools to decide upon more specific goals, content and methods. A national curriculum typically covers all subjects and courses offered in school education, whereas a **syllabus** provides more detailed information regarding the expectations for an individual school subject.
- **Educational standards** refer to descriptions of what students should know (content standards) and be able to do (performance standards) at different stages of the learning process. In some countries, standards are only available for the core subjects, such as literacy and mathematics, whereas in other countries they exist for a broad range of subjects. The standards may be set out in a separate document, or may be embedded in the curriculum.
- **Learning progressions** describe the way students typically move through learning in different subject areas. They can provide a roadmap for teachers to identify the set of skills and knowledge students must master on the way to becoming competent in more complex curriculum outcomes. Such learning progressions may be described in the curriculum or a separate document.

Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2) provide an overview of the references used for student assessment across OECD education systems at different levels of education. The references used in lower secondary education are listed in Table 4.1 below. As can be seen from the table, the majority of education systems use central or state curriculum goals as the main reference for student assessment. Central standards are also frequently used, either as the main reference for assessment or in addition to national curriculum goals. Central standards are used for at least one assessment type (internal summative assessment, national examinations and/or national assessments with no stakes for students) in Australia, Austria, the Flemish and French Communities of Belgium, Chile, the Czech Republic, Estonia, France, Hungary, Luxembourg, the Slovak Republic, Poland and Spain. Learning progressions are less frequently used across education systems. They serve as references for internal summative assessment and central assessments in Northern Ireland in the United Kingdom, for central assessments in Australia and Norway and for both national assessments and examinations in Denmark. Poland has specific examination standards based on the national core curriculum that serve as references for central examinations and Canada uses its own assessment framework for the development of the Pan-Canadian Assessment Program. In Denmark, there is no national curriculum in compulsory education, but binding national objectives were introduced in 2003 to serve as a reference for internal summative assessment. In Sweden, national knowledge requirements serve as the main reference for central assessments.

Table 4.1 References used in student assessment in lower secondary education – ISCED 2 (2012)

	Internal summative assessment	Central examinations	Central assessment
Central curriculum goals	Australia (pre-voc and voc), Austria, Belgium (Fl.), Belgium (Fr.), Finland, France, Iceland, Italy, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Slovenia, Spain, Portugal	Estonia (general), France, Ireland, Italy, Netherlands (pre-voc and voc), Norway, United Kingdom (Northern Ireland), Portugal[1]	Iceland, Israel, Italy, Korea, Mexico, Netherlands (pre-voc and voc), New Zealand, Slovak Republic (general), Slovenia
Central standards	Poland (plus curriculum decided at the school level), Slovak Republic	Belgium (Fr.) (general)	Austria, Belgium (Fl.), Belgium (Fr.) (general), Luxembourg
Central curriculum goals and standards	Australia (general), Chile, Czech Republic, Estonia, Hungary		Chile, Czech Republic, France, Spain
Learning progressions	United Kingdom (Northern Ireland)	Denmark	Australia, Denmark, Norway, United Kingdom (Northern Ireland)
Specific examination/ assessment standards		Poland (examination standards based on core curriculum) (general)	Canada (Pan-Canadian Assessment Framework), Finland (marking guidelines), Hungary (National Assessment Framework)
Binding national objectives/national knowledge requirements	Denmark		Sweden

Note: (1) National standards will be used as a reference for central examinations from 2013/14.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Providing clear goals and expectations for student learning to guide assessment

There are pronounced differences in the degree to which countries set system-wide expectations for student performance to guide teaching, learning and assessment across schools. While it is common for OECD countries to have system-wide curriculum frameworks that set objectives for student learning, the degree of prescription varies widely between countries.

Highly prescriptive central curricula can act as a legal trigger to promote evidence-based approaches and innovations regarding content and pedagogy and bring these within the reach of all schools and teachers (Elmore and Sykes, 1992; Westbury, 2007; in Kärkkäinen, 2012). But such curricula may not respond well to different local contexts and there may be a lack of ownership and commitment to change among teachers. On the other hand, highly decentralised curricula allow schools and teachers to experiment and develop curriculum innovations that are relevant to local contexts and may spread through horizontal networks of schools (Elmore and Sykes, 1992; Elmore, 1996; Darling-Hammond, 1998; Marsh and Willis, 2007; in Kärkkäinen, 2012).

Several education systems have adopted participatory approaches to developing curricula, where the national curriculum provides the core of overarching objectives whereas the more specific goals and curriculum content and assessment criteria are developed at the local and/or school level. This is the case, for example, in the Flemish Community of Belgium, the Czech Republic, Denmark, Finland, Norway and the Slovak Republic. Such curricular autonomy is intended to provide space for local interpretation and adaptation of goals. It is also expected to help strengthen local ownership of the teaching programme.

While it is important to keep curricula open so as to allow for teachers' professional judgements and innovations in the classroom, in some settings there are concerns about a lack of clarity regarding the specific goals to be achieved by all students in different subjects and year levels. There are often large variations in schools' capacity and expertise to implement effective local curricula and assessment approaches. This may lead to a lack of equivalence and fairness in educational opportunities for students across the country. In the absence of clear and specific system-wide objectives, teachers may find it difficult to develop concrete lesson plans, learning goals and assessment strategies that are in line with national expectations.

The introduction of more detailed national standards (or expectations or benchmarks or competence goals) for what should be taught, learned and assessed in schools has been debated and tried to varying extents in many countries over the last quarter century. Central standards are intended to provide consistency and coherence, especially in contexts where there is a high degree of local autonomy regarding the development of curricula, teaching programmes and assessment. Although it may appear straightforward to create statements of expected learning and levels of proficiency, experiences in different education systems have shown that it is not an easy task to identify clear and agreed standards and criteria (Looney, 2011b; Nusche et al., 2011a).

Research also reveals challenges in ensuring that the curriculum, standards, teaching and assessment are consistent (see also Chapter 3). The core logic of standards-based systems rests upon the alignment of these key elements. If the assessments do not well match the curriculum and the standards, then assessment results have little value in judging how well students are learning. This, in turn, will make it difficult to diagnose and respond to student or school needs. Hence, policy needs to give considerable attention to sound strategies that assess student performance in relation to the curriculum and the standards. In addition, teacher education and professional development also need to be aligned to overarching curriculum goals and standards. Where curricula and assessments have been reformed to focus on key competencies, it is important to ensure that teaching and learning approaches are changed accordingly to provide opportunities for students to indeed acquire these competencies.

Procedures

This second section describes the procedures and methodologies used for student assessment across countries. This includes the scope of assessment, i.e. the areas of learning that are covered by the assessment as well as the key procedural features of student assessment across countries, i.e. the mix of instruments used in specific student assessment systems; the format of assessments; and the use of ICT in assessment. It also reviews ways in which the design of assessments can enhance or threaten fairness and equity in education.

Aspects assessed

A comprehensive body of research has described the different dimensions of learning that might be captured in assessment. Traditionally, classifications of learning outcomes have been based on the distinction between cognitive and non-cognitive learning, but more recently the concept of competencies which encompasses both cognitive and non-cognitive learning has become widely accepted. Box 4.6 provides an overview of different types of learning outcomes that education systems may seek to achieve.

While there is a strong ambition across OECD countries to focus school systems increasingly on the development of complex competencies, the OECD Country Reviews found a concern across countries that assessment systems might be lagging behind such competency-based curricula. Both standardised and teacher-based assessment often remained more traditional and focussed on isolated knowledge and the application of basic skills. While no directly comparable information is available regarding the scope of student assessment across countries, information regarding the subjects assessed and the assessment instruments used provides some indications about the scope of learning that is typically captured in current approaches to student assessment.

Box 4.6 Classification of learning outcomes: A brief overview

Cognitive learning refers to the construction of thought processes. Most current classifications of cognitive learning are inspired by the "taxonomy of educational objectives" developed by Bloom in the 1950s. Bloom divided the cognitive domain into six categories of increasing complexity: knowledge, comprehension, application, analysis, synthesis and evaluation (Bloom, 1956). Bloom's taxonomy has since been further developed, modified and refined by many authors (e.g. Gagné, 1977; Klemp, 1977; Kolb, 1981; Eraut, 1990, Marzano, 2001). Most authors of the assessment literature have retained two major categories of cognitive learning: the *acquisition of knowledge* and the *development of skills* (Posner, 1992). The acquisition of knowledge involves the recognition or recall of ideas, materials and phenomena, whereas the development of skills involves more complex processes of analysis, synthesis and evaluation. Sometimes, within the "skills" category, an additional distinction is made between *basic skills*, such as the routine application of well-practiced knowledge and skills, and *higher-order skills,* such as problem solving and the transfer of existing knowledge and skills to new situations (Crooks, 1988).

Non-cognitive learning refers to psychosocial development and the development of attitudes and values. Psychosocial development can include self-development (e.g. self-esteem, identity development) as well as relationships with other people and institutions (e.g. interpersonal and intercultural skills). Attitudes and values are closely related. *Attitudes* can be defined as beliefs focused on a specific object whereas *values* refer to more generalised standards that transcend attitudes (Pascarella and Terenzini, 2005). Attitudinal and value outcomes of learning can include development of social responsibility, motivation for learning and understanding of diversity (Volkwein, 2003).

Competency development refers to a broader model of learning which comprises a complex integration of knowledge, skills, attitudes, values and action (Baartman et al., 2006). A competency can be defined as "the ability to meet demands or carry out a task successfully and consists of both cognitive and non-cognitive dimensions" (Rychen, 2004). Competencies are complex ability constructs that are applied in real-life contexts and have to be acquired by learning. The concept of competency is based on the idea that the traditional taxonomies of learning cannot adequately capture the nature of learning that unites different types of knowledge, skills and attitudes into real expertise (Ewell, 2005). In recent years, many OECD countries have reformed their education goals and/or curricula to emphasise the importance of "key competencies" or "21st century competencies". While the definition of such "key competencies" varies across countries, they typically include elements of learning to learn skills, information processing, communication, teamwork, critical analysis and creativity.

Source: Reproduced from Nusche (2008).

Subjects assessed

In all OECD countries, schools are expected to monitor their educational quality and assess the extent to which students acquire the knowledge, skills and competencies outlined in the curriculum or performance standards. Typically, students are assessed by their teachers in all subjects that are part of the curriculum. Assessments tend to be continuous throughout the year and are generally complemented by more high stakes teacher-based or external examinations at certain key stages in the school year or cycle.

While it is generally expected that all subjects are given some attention in teacher-based assessment, standardised assessments and examinations – where they exist – tend to focus on a few priority subjects. The subjects assessed in central examinations vary across education systems. As shown in Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2), the subjects most frequently assessed in OECD countries are the language of instruction and mathematics, at all levels of education. But in many education systems, a range of other subjects are also assessed.

Subjects covered in central examinations

Only a few OECD education systems have central examinations in place in primary education (Figure 4.3). The French Community of Belgium, Portugal and the United States assess all primary school students in mathematics and the language of instruction. In addition, both the French Community of Belgium and the United States assess students in natural sciences, and the French Community of Belgium also assesses students in social sciences. In Turkey, national examinations exist in primary education in Year levels 6, 7 and 8, but they are not compulsory and students are given a choice of subjects for examination (OECD, 2012a).

Fourteen OECD education systems implement central examinations in lower secondary education. Thirteen of them are depicted in Table 4.2 below (Canada is not included in Table 4.2 because the characteristics of examinations vary across provinces). As shown in the table, all education systems except Norway and Scotland in the United Kingdom have some examination subjects that are compulsory for all students at this level. The subjects most frequently examined in a compulsory manner are the language of instruction (11 systems) and mathematics (10 systems). Other frequently examined compulsory subjects include the sciences and modern foreign languages. In Scotland, there are no compulsory examination subjects, but students choose among a range of possible subjects. In four countries (Denmark, Estonia, Ireland and the Netherlands), students can choose among a range of optional subjects in addition to a number of compulsory subjects. Norway and Denmark use an approach where students are randomly selected to sit an examination in a given subject. In Norway, students in Year 10 are randomly sampled to sit a centrally administered written examination in one subject among Norwegian, mathematics and English. In Denmark, in addition to a range of compulsory and optional examination subjects, students are randomly sampled to sit examinations in social sciences and religion.

Twenty-five OECD education systems have central examinations in upper secondary education, and most of them have at least one compulsory examination subject that all students have to take (OECD, 2012a, complemented with information collected from countries participating in the OECD Review). The language of instruction is the most frequent compulsory examination subject, followed by mathematics and modern foreign languages. Examinations in upper secondary education are characterised by a wider range

of options for students. While in lower secondary education less than half of the countries have optional examination subjects for students, in upper secondary education this is the case for over two-thirds of the countries. As in lower secondary education, Denmark and Norway use a sampling approach, where students are randomly selected for examinations in most subjects (OECD, 2012a).

Table 4.2 Characteristics of central examinations in lower secondary education – ISCED 2 (2011-12)

Country	Programme type	Are central examinations compulsory?	Which year levels are assessed?	What subjects are assessed?		
				All students	Students choose	Sample
Belgium (Fr.)	General only	Yes, for public schools only (mandated at the Community level)	8	M, L		
Denmark	All programmes	Yes, for public schools only*	9	M, L, S, FL	A, V	FL, S, SS, R
Estonia	General only	Yes, for public and government-dependent private schools only	9	M, L	S, SS, FL	
France	All programmes	Yes, for all schools	9	M, L, S, SS, FL, T, A, R		
Germany	General only	Yes, for all schools	9, 10	M, L, FL		
Ireland	All programmes	Yes, for public and government-dependent private schools only	11	M, L, S, SS	A, R, V, O	
Italy	All programmes	Yes, for all schools	8	M, L, S, FL, A		
Netherlands	All programmes	Yes, for all schools	12	L, FL, V	M, S, SS	
Norway	All programmes	Yes, for all schools	10			M, L, FL
Poland	General only	Yes, for all schools	9	M, L, S, SS, FL		
Portugal	General only	Yes, for all schools	9	M, L		
United Kingdom (Scotland)	All programmes	No (but 100% administer them)	11		M, L, S, SS, FL, T, A, R, V	
United States	All programmes	Yes, for public schools only (mandated at the state level)	6, 7, 8, 9	M, L, S		

Notes: Subjects assessed: M: mathematics, L: national language or language of instruction, S: science, SS: social studies, FL: modern foreign language, T: technology, A: arts, R: religion, V: practical and vocational skills, O: other.

All students: all students take the test; Students choose: students can choose to take the test in this subject; Sample: sample or selection of students take the test in this subject.

*) Denmark: 95% of government-dependent private schools also administer the examination.

Source: OECD (2012a), *Education at a Glance 2012*, adjusted and complemented with information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Assessment criteria and scoring rubrics

In order to help markers make a judgement about the extent to which standards have been met, many education systems have developed additional documents that detail and illustrate different levels of proficiency. Such guidelines can help create a shared understanding regarding what constitutes adequate, good and excellent performance in

different subject areas. They can also provide information regarding the level of performance required for a particular mark. A range of tools can contribute to facilitate the assessment of student work against national curricula, standards or learning progressions:

- **Performance criteria** refer to guidelines, rules or principles by which student performances or products are judged. They describe what to look for in student performances or products to judge quality.
- **Rubrics** refer to scoring tools containing performance criteria and a performance scale with all score points described and defined.
- **Exemplars** refer to examples of actual products or performances to illustrate the various score points on a scale.

Criteria or rubrics may be part of the national curriculum or syllabi, or be provided in a separate document. In Finland, for example, descriptions of good performance and criteria for final assessment are determined within the core curriculum. While the marking scale is from 4 (weak) to 10 (excellent), the national core curriculum only defines criteria for the mark 8 (good) for each subject at transition points. These criteria are intended to help teachers focus assessment on core knowledge and competencies and understand the meaning of "good learning" at different stages. In Ireland, assessment criteria have been developed in secondary education for each subject that is taken in national examinations (the Junior Certificate and the Leaving Certificate). These criteria are closely linked with the syllabus for each subject and are usually an integral part of the discussion during the development of a new syllabus or course. In the Netherlands, benchmarks in literacy and arithmetic were introduced in 2010 to set a common framework of expectations and learning trajectories across all educational tracks in these two core subjects. Each benchmark provides a general description, a description of the tasks that students should be able to perform and the criteria these tasks have to meet, for two levels: a fundamental level that all students should meet and a more advanced level for gifted students (Scheerens et al., 2012).

Information from the education systems participating in the OECD Review indicates that the development of rubrics that detail assessment criteria for systematic and widespread use are not yet common place across countries. In Denmark, for example, the Common Objectives provide goals that articulate the knowledge and skills that "teaching should lead towards" but appear to lack performance standards that describe concrete learning outcomes that students should achieve and criteria for assessing these (Shewbridge et al., 2011). In Norway, school owners have expressed concern about the lack of standards concerning the competencies required for a particular mark and the potentially resulting unfairness in teacher marking (Norwegian Directorate for Education and Training, 2011, in Nusche et al., 2011a). In Sweden, new curricula were introduced at all ISCED levels in 2011 to provide clearer goals and guidance for teachers regarding expected learning outcomes for different subjects and year levels. This was in response to concerns that the learning goals provided by the previous core curriculum remained too vague and led to inequities in teacher marking (Nusche et al., 2011b). In the Slovak Republic, while national performance standards were recently introduced, the level of detail of the criteria to measure achievement of these standards varies among subjects (Hajdúková et al., forthcoming). In the French Community of Belgium, the Inspectorate noted large differences in the required performance levels across schools, despite the existence of common competency references (Blondin and Giot, 2011).

In many settings, assessment criteria are defined at the teacher or school level. In the Czech Republic, for example, criteria are defined by each school and reflected in the respective School Education Programme. As a result, even though all schools may use the same marking scale, they have different marking criteria (Santiago et al., 2012b). As in many other countries, it is not necessarily common practice for teachers to specify assessment criteria in full detail and to inform students of them in advance. In Portugal, there is a strong national focus on developing assessment criteria. At the same time, while national guidelines exist for the development of criteria, schools have some flexibility, for example in the weightings they can assign to different components. Within the nationally defined guidelines, schools are autonomous to specify and publish their own assessment criteria. These criteria are shared with students and guardians and are used to support internal summative assessment as well as decisions about student progress or additional support required (Santiago et al., 2012a).

Qualification frameworks

At the secondary level, several education systems have introduced qualifications frameworks to provide clarity to stakeholders within and outside educational institutions regarding the academic and vocational qualifications students can obtain and the associated competencies they need to acquire. A qualifications framework can be defined as a rank order of qualification levels, allowing each qualification to be assigned to a specific rank. It classifies qualifications according to a set of criteria for levels of learning achieved (OECD, 2007, 2010b). In Europe, the development of the European Qualifications Framework has encouraged the development of national frameworks, in line with the European framework. Education systems such as the Flemish Community of Belgium, Hungary, Ireland, Spain and the United Kingdom have recently introduced such frameworks (OECD, 2010b).

Qualifications frameworks vary across education systems in the level of prescription they provide. In some countries, including the United Kingdom, New Zealand and South Africa, they have a strong regulatory function with common rules across all qualifications. In other education systems, such as Australia and Scotland in the United Kingdom, they are more loosely defined. In these cases, the framework provides more of a map of qualifications available and allows room for differences at the local level (OECD, 2010b). While the implementation of qualifications frameworks is often challenging, there are important benefits in terms of providing clear references regarding the level of competencies that should be associated with different qualifications and how the different qualifications relate to each other. Such transparency in progression pathways at the secondary level can facilitate students' progression in education, while at the same time acting as a quality assurance mechanism for qualifications offered by different providers (OECD, 2010b).

Assessment instruments and approaches

Much has been written about the advantages and disadvantages of different assessment instruments and the type of learning that different instruments can capture. Looney (2011b) describes four broad groups of assessment approaches frequently used in standardised assessments:

- *Multiple-choice assessments* present students with a set of alternative answers. Students must choose one answer from this set. They are machine-scored and provide comparable data on student results. Well-designed multiple-choice items

can assess higher-order knowledge, but they cannot assess broader skills and competencies such as the ability to develop an argument or communicate effectively. Poorly designed multiple-choice tests may be prone to measurement error, for example when students misinterpret questions or make random guesses.

- *Adaptive assessments* are computer based and, as implied by their name, adapt questions to the test-taker. Students who answer questions correctly are directed to a more difficult question and those answering incorrectly receive an easier question. Such adaptive tests can provide more fine-grained information on student performance than traditional multiple-choice tests. However, since not all students respond to the same questions, it is not possible to compare student performance. Also, adaptive tests require a very high number of test questions, which contributes to higher development costs.
- *Performance assessments* are a range of alternative assessment approaches developed partly as a reaction to the widespread use of multiple-choice tests in some countries. In other countries, there is a long tradition of performance assessments and only very limited use of multiple-choice tests. The main characteristic of performance assessment is the intention to assess a range of integrated knowledge, skills and attitudes by asking students to perform a task rather than to provide a correct answer. They may include tasks such as essays, oral presentations, portfolios, experiments and group work. Such assessments tend to capture a broader range of curriculum goals and may be more effective at measuring complex competencies. However, there are concerns regarding the reliability of these assessments. Since scores are typically awarded by human raters, there may be variability in their judgements. Such assessments may also be more expensive to administer and score.
- *Computer-based performance assessments* can potentially assess complex performances by using information and communication technologies (ICT), such as simulations, interactivity and constructed response formats. In some countries, especially in the United States, research has been conducted to develop increasingly sophisticated ICT programmes that are able to score open-ended performances. Such assessments are not yet widely used and still in the early stages of development. They may help address concerns related to the reliability of human raters on the one hand and related to the validity of multiple-choice assessments on the other.

Instruments used in standardised assessments and examinations

OECD education systems use a range of different formats and approaches for standardised central assessments of student learning. Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2) provide information on the assessment formats applied in standardised assessment in mathematics and the language of instruction. As is typical for standardised assessment, the vast majority of education systems use written formats for their central assessments and examinations. The instruments used do not differ very much between assessments in mathematics and assessments in the language of instruction within education systems. This section provides information on assessment formats used at the lower secondary level (ISCED 2).

In standardised central examinations that have formal consequences for individual students, the most frequently used assessment formats are open-ended written tasks. Multiple-choice items are also frequently used, especially in examinations in the language

of instruction. A few education systems also use closed-ended short-answer formats. Denmark uses more performance-based formats, namely oral presentations, oral questions and answers and project presentations as part of the standardised national examinations (in the language of instruction only). In Italy, the school-based part of the examination also includes oral questions and answers and oral presentations. Estonia uses performance-based tasks as part of the mathematics examination. In the case of central examinations, the strong focus on written tasks appears to be related to concerns about reliability and fairness in assessment and marking of standardised assessments. In many countries, these centrally designed standardised components are complemented by non-standardised parts of the examinations that are locally designed and marked. While there are limits to what any centrally administered standardised assessment can assess, it is often expected that the assessment of a broader range of skills and competencies happens in such local assessments.

In central assessments that do not have formal consequences for students, multiple-choice tests are by far the most frequently used assessment format in both mathematics and the language of instruction. Closed-format short-answer questions and open-ended writing tasks/open calculations are also frequently applied. Many countries use a mix of these three formats. Only a few countries including Austria (language of instruction only), Iceland, New Zealand and Sweden use oral question and answer formats. Only Austria (language of instruction only), Finland (language of instruction only) and Sweden require students to perform oral presentations as part of the national assessments. The use of performance tasks is also limited; such tasks are applied in central assessments in the Flemish Community of Belgium, the French Community of Belgium (general programmes only), and New Zealand. The limited use of oral and performance-based tasks in central assessments might be explained by the fact that such assessments serve primarily for monitoring and comparing student results across regions and schools and the purpose is to obtain highly reliable and easily comparable scores. The administration of tasks that require one-to-one assessment situations would also be more costly and time-intensive than written tests.

Moving towards more innovative forms of standardised assessment

While central assessment systems tend to rely predominantly on traditional paper-based assessment formats, there are several interesting examples of assessments that attempt to capture a wider range of competencies. Innovative performance-based assessment formats offer significant potential to signal the learning goals that are valued in the education system. They can also model to teachers the next generation of assessment formats that can measure and support broader learning.

To help teachers use innovative forms of assessment, some countries have developed specific assessment instruments that are made available for teachers to use when they see fit. Several countries have developed on-demand assessments, where teachers can draw from a central bank of assessment tasks and ask students to take the assessment when they consider that they are ready. This gives teachers control of the timing of assessment. In Australia, for example, the government's Online Diagnostic Tools initiative provides teachers with access to online resources that can help assess student progress and provide links to relevant digital learning resources to help them improve. Assessment instruments may also be developed by other actors in the education system. In the Flemish Community of Belgium, for example, several specific evaluation instruments have been developed by various institutions to help teachers assess the non-cognitive performances of their students (Box 4.7).

Box 4.7 Assessment formats that measure broader competencies

In the **Flemish Community of Belgium**, a range of institutions including educational centres, academic institutes and umbrella organisations, have developed tools that teachers can draw on to assess non-cognitive aspects of learning, such as attitudes, well-being and involvement. Some of these tools are used quite widely while others are applied rather sporadically and in a limited number of schools or education forms. The most commonly used tool for assessment of non-cognitive performance is the SAM-scale (Scale for Attitude Measurement; *Schaal voor AttitudeMeting*)[7]. The SAM-scale measures students' attitudes and has been developed to assess, guide, stimulate and orientate students. By means of this tool a teacher can determine to what extent a pupil scores high or low for certain attitudes, e.g. flexibility, diligence and responsibility.

In **Finland**, "learning to learn" skills are considered to be central to each student's development. These are actively promoted as core elements in achieving lifelong learning and include a student's capacity for independent and self-motivated learning, problem-solving and the ability to evaluate his/her own learning and related strategies. There is a clear pedagogical goal in all compulsory education subjects for students to develop "learning to learn" skills. To evaluate and promote the importance of such skills, national sample assessments were developed by the Centre of Educational Assessment at the University of Helsinki to evaluate "learning to learn" skills in Years 3, 6 and 9 of compulsory education.

In **New Zealand** primary schools, progress towards the achievement of national curriculum goals is measured via the National Education Monitoring Project (NEMP). NEMP is designed to be as well aligned as possible with the curriculum by incorporating competency and value elements. Many of the NEMP assessment tasks are performance-based, requiring students to transfer learning to authentic close-to-real life situations. There are different assessment situations including one-to-one interviews, work stations and teamwork. As the assessment does not carry high stakes for students it is particularly important that tasks are meaningful and enjoyable to them. The assessment provides rich information on the processes used by students to solve problems or conduct experiments. Most assessment tasks are carried out orally so as to analyse what students can do without the interference of reading and writing skills. Some of the tasks are videotaped to allow for an in-depth analysis of student responses and interaction with teachers. NEMP also assesses students' cross-curricular skills, and attitudes towards the learning areas being assessed. Students' enjoyment of particular assessment tasks is also surveyed. For instance, 82.5% are reported as enjoying hands-on tasks, versus around 50% for paper and pencil tests. While NEMP is designed for system monitoring, examples of previous assessment tasks are available for teachers and may be used in the classroom. This can help teachers estimate how their own group of students compares to national assessment results.

Sources: Flemish Ministry of Education and Training (2010); Finnish Ministry of Education and Culture (forthcoming); Nusche et al. (2012)

Other countries have developed sample-based surveys that cover larger parts of the curriculum including non-cognitive elements (see Box 4.7 for examples from Finland and New Zealand). While these primarily aim to provide information at the education system level, they also give important insights to participating students and schools regarding valued key competencies. In addition, where the assessment tasks of previous years are published, these can serve as example tasks for formative assessment in all schools. In New Zealand, for example, such use of previous assessment tasks is encouraged by the education authorities.

As will be discussed further below, developments in information and communication technology (ICT) also offer new avenues for assessing broader competencies on a larger scale. Of course, ICT-based technology is not a prerequisite for large-scale assessment of students' open-ended performance, but they may make such assessments considerably more affordable.

Instruments used in teacher-based assessment

In the education systems participating in the OECD Review, teachers are largely autonomous in the choice of internal assessment instruments within the limits of the local or national curriculum. Typically, there are no central regulations regarding the particular assessment instruments to be used in measuring the achievement of learning goals. Given the diversity of approaches applied by teachers in different subjects, year levels, regions, schools and classrooms, it is difficult to characterise and compare "typical" assessment formats used within a system. However, the information collected through Country Background Reports and OECD Country Reviews can give some initial indications and observations regarding internal assessment trends within countries.

It appears that, traditionally, teachers in all education systems have relied as a minimum on written assessments and examinations and certain oral assessments when measuring student performance in core subjects. Written assessments may take many different forms, such as multiple-choice tests, quizzes, dictations, exercises, short and long constructed response tasks and essay questions, while oral examinations typically take the form of question and answer exercises. The traditional reliance on rather formal assessment formats may be explained by the way teachers' assessment role has long been conceptualised in many countries. In many settings, the understanding of assessment has been one of summative intention, serving primarily to provide reliable and comparable results and rankings of students in order to make decisions about placement, advancement and selection. In many countries, the checking of homework assignments and observation of classroom participation may also be part of classroom-based assessment. Depending on the subject, particularly for subjects requiring more practical evidence of student performance such as physical education, music, arts and sciences, it is reported that teachers have traditionally relied also on broader formats such as practical demonstrations, performances and experiments.

Teachers across education systems participating in the OECD Review continue to draw on the abovementioned assessment approaches. At the same time, in many settings, teachers are going beyond traditional assessments and report about their use of a much broader mix of assessment approaches including more sophisticated assessment types aiming to capture a broader range of integrated knowledge and skills. In most countries where OECD Country Reviews were organised, teachers reported using various forms of alternative assessment, including products (e.g. written essays; laboratory reports), performance (role plays, experiments, presentations) and portfolios. Such a range of assessments using *product, portfolio and performance*, are also known as "3P assessments" and reflect an emphasis on assessing "higher-order" thinking skills in authentic ways (Madaus and O'Dwyer, 1999; Stiggins, 1987). In several settings, teachers also reported relying more on observing and recording student achievement as it occurred and mapping progress through the collection of student work samples over time.

Promoting the use of innovative assessments by teachers

To support a balanced approach to teacher-based assessment, several education systems have recently launched central initiatives to help establish a new culture of assessment focussed on more performance-based approaches to measuring student learning. In France, for example, personalised competency monitoring systems were introduced for teachers to record their students' progress in acquiring core knowledge and competencies (Box 4.8).

Box 4.8 Continuous teacher-based assessment of student competencies

In **France**, the 2005 orientation law highlights the requirement for schools to ensure that all students achieve a common core (*socle commun*) of knowledge and competencies necessary for their success in school as well as their personal and professional future. In this context, teachers are required to assess their students' progress towards the common core of competencies using the students' Personal Competency Booklet (*Livret Personnel de Compétences, LPC*) throughout their compulsory schooling. A range of items that students should have acquired at key stages of education are defined through the legislation, and teachers attest students' mastery of these items on the Personal Competency Booklet. Items are validated by classroom teachers in primary education and by the class councils in lower secondary education, with the school leader approving the overall mastery of the common core of competencies at the end of lower secondary education. While it is not obligatory for students to validate each item (this is left to the discretion of teachers), it is mandatory for them to validate each of the seven key competencies defined in the 2005 orientation law. To support this assessment process, all teachers have received guidance material defining the different competencies students are expected to achieve and providing a number of assessment tools and reference criteria to undertake the assessment.

Source: Dos Santos and Rakocevic (2012).

In other systems, support for broader assessment formats takes the form of central frameworks or guidelines for assessment suggesting a range of different assessment approaches to teachers so as to ensure that different types of learning are given adequate attention across the curriculum.

In Austria, for example, a new secondary school type (the "New Secondary School") was launched in 2008 as an inclusive school to avoid the early tracking of students into different school tracks after Year 4. This attempt at a structural reform is accompanied by pedagogical reforms such as new ways of individualisation and differentiation of instruction, social learning and integration of students with special needs. In line with the focus on promoting innovative pedagogical approaches, the new secondary schools are required to use more formative forms of assessment and to focus on the assessment of performances, through, for example, self-observation, self-assessments and portfolios (Specht and Sobanski, 2012).

In the Flemish Community of Belgium, the central education authorities are promoting a shift towards a "broad assessment culture" which includes a focus on formative assessment and new assessment approaches. It implies the use of "alternative" (compared to tests) assessment approaches including observation, portfolios, reflection sheets and self- and peer-assessment. It is also being highlighted that it is more important to report on student progress compared to previous achievements rather than on absolute performance. While it takes time to bring about such a culture shift systematically across all schools, progress has been made in promoting alternative forms of assessment, in particular in primary education, vocational education and special needs education (Flemish Ministry of Education and Training, 2010).

In Ireland, the Primary School Curriculum provides for the use of assessment approaches such as teacher observation, teacher-designed tasks and tests, work samples, portfolios and projects, curriculum profiles, diagnostic tests and standardised tests. The National Council on Curriculum and Assessment (NCCA)'s guidelines on assessment present these and other assessment approaches as a continuum of approaches moving from those that are pupil-led such as self-assessment and conferencing to those that are more teacher-led such as teacher observation, teacher-designed tasks and tests, and standardised testing. Assessment approaches such as portfolio assessment, concept

mapping and questioning appear towards the middle of such a continuum. In secondary schools, informal methods of assessment include classroom tests administered at the end of topics or segments of the syllabus to assess students' knowledge and mastery of particular skills, informal observation by the teacher, evaluation of homework, and informal analysis of students' language and social development (Irish Department of Education and Skills, 2012).

In Mexico, the national curriculum (Study Plan) states that rubrics, checklists, registries of observations, written pieces of work, team projects, conceptual maps, portfolios and written and oral tests should be used. It also requires that students should be frequently involved in self-assessment and peer assessment activities (Santiago et al., 2012c).

In New Zealand, also, the importance of using multiple sources of evidence for effective assessment is emphasised at both primary and secondary levels. Schools are required to use a range of assessment practices to measure students' progress and achievement in relation to the national curriculum and standards. Thereby, they are encouraged to use a diversity of approaches in school-based assessment. The focus on broad assessment is further emphasised by the reliance of National Standards on overall teacher judgements (OTJ) rather than a national standardised test (Nusche et al., 2012).

In Slovenia, the principles for assessment and examinations are specified in Rules on Examination and Assessment of Knowledge for the different levels of education. Among other things these principles require teachers to use a variety of forms and methods of verification and assessment of knowledge; to take into account the ability to analyse and interpret creativity and the ability to use knowledge; to allow students a critical reflection and insight into the acquired knowledge and to contribute to the democratisation of relations between students and teachers. In primary education, student achievement is assessed continuously in written, oral, artistic, technical and practical forms, as well as through tests set by teachers. In upper secondary education, teachers assess students' oral answers, written tests, visual arts and other works, seminars, projects and presentations (Brejc et al., 2011).

In Singapore, the "thinking schools, learning nation" initiative was introduced in 1997. The explicit focus of the related reforms was on developing a creative and critical thinking culture within schools, developing an inquiry orientation among teachers and students and assessing students on these competencies. These initiatives were also linked to a commitment to integrating technology into all aspects of education. New subjects, such as "project work" and "knowledge and inquiry" were introduced, along with requirements to use specific performance-based assessments in these subjects. Many courses of the reformed curriculum include applied examination elements that allow students to show their problem-solving and thinking skills (Darling-Hammond, 2010) (Box 4.9).

In some countries, assessment approaches developed in the vocational education and training (VET) sector or in second-chance education seemed to be ahead of the general sector in terms of paying attention to wider competencies and making efforts to assess these through sophisticated assessment approaches.

In Finland, for example, all vocational qualifications include skills demonstrations, which form part of student assessment. The demonstrations take place in a practical and authentic work situation, where the student shows how well he or she has attained the objectives and aims set for vocational competence in the Core Curriculum. In 2004, the Finnish National Board of Education commissioned a review of the effect of skills

demonstrations on the quality of vocational education and training. The study concluded that skills demonstrations have had several positive effects. The system was found to help assure the level of students' learning, respond to the needs of the labour market and enhance the quality of VET education, because the feedback received from professionals could be used in the further development of training and teaching. Overall, the review indicated that the skills demonstrations had a positive effect on students' motivation and aptitude to learn and increased the value of VET in the labour market (Finnish Ministry of Education and Culture, forthcoming).

Box 4.9 Singapore: Creating "thinking schools"

In Singapore, recent reforms have changed the curriculum and assessment system to make it more explicitly focused on creativity and independent problem solving. Curriculum and assessment guidelines encourage teachers to use a variety of assessment formats. The Ministry has developed support tools for teachers such as:

- The Strategies for Active and Independent Learning (SAIL) which aim to support learner-centred project work and provide assessment rubrics to clarify learning expectations. All schools have received training for using these tools.
- The Ministry's 2004 Assessment Guides for primary and lower secondary mathematics which contain resources, tools and ideas to help teachers incorporate strategies such as mathematical investigations, journal writing, classroom observation, self-assessment and portfolio assessment into the classroom. The Institute of Education has held a variety of workshops to support learning about the new assessments and integrated the new strategies into teacher development programs.

In addition, Project Work was introduced as an interdisciplinary subject that requires students to draw knowledge and apply skills from across different subject domains. The requirements for project tasks are centrally set by the Singapore Examinations and Assessment Board. The tasks are defined in a broad way so as to allow students to carry out a project of their interest while meeting the following task requirements: *(i)* it must foster collaborative learning through group work, *(ii)* every student must make an oral presentation; and *(iii)* both product and process are assessed and there are three components to assess them: a written report, an oral presentation and a group project file to which each group member submits three documents related to snapshots of the processes involved in the project. About 12 000 students complete this task annually. Assessment is school-based and criterion-referenced. While task setting, conditions, assessment criteria, achievement standards and marking processes are externally specified by the Singapore Examinations and Assessment Board (SEAB), the assessment of all three components of Project Work is carried out by classroom teachers, using a set of assessment criteria provided by the board. All schools are given exemplar material that illustrates the expected marking standards. The Board provides training for assessors and internal moderators. Like all other assessments, the marking is both internally and externally moderated.

In a recent paper, Koh et al. (2011) analyse the Singaporean experience and highlight the need to bring about fundamental changes of school cultures and teacher and student dispositions in order to make these assessment reforms work in practice. They find that despite policy change teachers often remained reliant on conventional assessment and were sometimes ill-prepared to implement authentic assessment. In a two-year empirical study on teacher professional development in authentic assessment and use of rubrics, they found that teachers were able to improve the quality of classroom assessment after participating in certain types of professional development. Not surprisingly, active and collective participation in ongoing, sustained professional development was more effective than ad hoc or 1-2 day workshops in building teachers authentic assessment capacity.

Sources: Darling-Hammond (2010); Koh et al. (2011).

In Portugal, with the so-called New Opportunities programme, there have been a number of initiatives to extend the educational provision in schools to students who may have left school, or may be at risk of leaving school, and to adults who might not have completed compulsory education. These new programmes have been accompanied by the development of approaches to assessment focused on motivating students, giving high-quality feedback, and including the active participation of learners in the assessment process. A key feature of these arrangements and approaches is their location close to the learning process and to the learner. Thus, assessment tends to occur immediately after the completion of a module or portion of a course rather than at the end of a year or cycle. The use of approaches beyond written tests, such as a performance assessment, puts the learner and learning at the centre of the assessment process. The need to use assessment to motivate learners to learn, rather than to simply engage with the assessment or test, is acknowledged as fundamental by those working in this sector (Santiago et al., 2012a).

While innovative assessment of practical skills is essential in vocational education and training, OECD (2010b) argues that such assessment should be embedded within a standardised central assessment framework. The advantage of having a standardised framework for assessment in the VET sector is to ensure that all those with a qualification have the same mix of competencies, in particular where there is substantial variation among individual VET institutions and companies offering apprenticeships. There are a range of potential approaches to ensuring common standards in assessment in VET, such as periodic inspections of VET institutions, inspections of examination bodies, random evaluation of student performance, self-evaluation of providers, peer reviews, clear central guidelines for locally developed examinations and availability of a range of standardised assessment tools (OECD, 2010b).

Ensuring consistency of marking through moderation

A key strategy to increase the reliability of assessment and marking across different settings is to systematically implement moderation procedures. Moderation refers to a set of approaches that aim to ensure the quality and comparability of assessment judgement. It may involve teachers cross-marking each other's assessments or discussing student performance in groups, or a competent external organisation systematically checking school-based marking.

While in many settings, moderation occurs informally within and between schools and may not be documented, some education systems have introduced systematic arrangements for moderation. This is particularly the case in education systems where centrally developed examinations with high stakes for students are corrected and marked locally by teachers. In the French Community of Belgium, schools are in charge of marking their own students' examinations and they decide autonomously whether students' examinations are corrected by their own teacher, another teacher or a group of teachers. Guidance materials are available to support schools in this task. In France, while teachers examine their own students through continuous classroom assessment, teachers from another school are responsible for marking written examinations leading to diplomas or certification. In Denmark, centrally appointed external examiners correct examination papers and are assisted through national guidance materials such as performance criteria, exemplars, rubrics and keys. There is also moderation of marking by external examiners who attend oral examinations. In the Netherlands, examinations are corrected by the students' own teacher and moderated by a teacher from another school using a central scoring protocol. The school boards are responsible for the proper

handling of the procedures. In case of disagreement, external moderation by a competent body is provided (Tables 4.A1.4, 4.A2.4a and 4.A2.4b, Annexes 4.A1 and 4.A2).

In Australia, while each state and territory has its own system and own set of procedures for developing and approving courses in secondary education, most combine student performance on external exams at the end of Year 12 with moderated, teacher judgements of coursework performance to arrive at scores for senior secondary certificates and high school completion. In Queensland, the examination system is school-determined and based, but achievement standards and scoring are externally moderated. Moderation processes for the Senior Certificate (Year 12) involve subject-based panels of expert teachers providing advice to schools on the quality of their assessment programme and their judgements of quality of student performance based on sample portfolios. The system involves follow-up with schools where panels identify issues regarding assessment and standards. There is negotiation of the final results to be recorded on the Senior Certificate (Sebba and Maxwell, 2005 in Santiago et al., 2011). Similarly, procedures adopted by educational jurisdictions and particular schools for moderating internal summative teacher judgements (so-called A-E ratings) also facilitate common understanding of year level proficiency standards and foster the development of professional learning communities that can provide crucial support for improving opportunities for student learning and building teacher capacity (Santiago et al., 2011).

In New Zealand, an external moderation system is also in place to ensure the dependability of internal assessments in Years 11-13. The New Zealand Qualifications Authority directly checks the quality of internal assessment through a sampling approach. Schools are required to submit 10% of internally assessed student work for NZQA moderation to make sure the assessment is appropriately aligned with standards. The moderation process does not affect the marks assigned to assessment samples by teachers, but is intended to provide feedback to teachers and to inform future assessment policy development at the system level (Nusche et al., 2012).

Use of technology in assessment

In recent years, the potential of information and communication technologies (ICT) to influence and shape assessment approaches has been increasingly recognised across OECD countries. Binkley et al. (2010) describe two key strategies regarding the use of ICT in assessment. First, the "migratory" strategy refers to the use of ICT to deliver traditional assessment formats more effectively and efficiently. Second, the "transformative" strategy refers to the use of ICT to change the way competencies are assessed and develop formats that facilitate the assessment of competencies that have been difficult to capture with traditional assessment formats (Binkley et al., 2010; Ripley, 2009; European Commission, 2011a).

Information collected from education systems participating in the OECD Review shows that the use of ICT for assessment has not yet become common practice internationally. In the few systems where technology is used for standardised central assessments or examinations, it is mostly done in a "migratory" perspective. Some countries, including Australia, the Czech Republic and New Zealand use technology in national examinations for data management purposes, such as data sheet scanning, inputting marks or results management, while examinations remain paper-based. In a few countries, such as Slovenia, computer-based technology is used for students with special educational needs. Some education systems, namely the Flemish Community of Belgium, the Czech Republic, Luxembourg (ISCED 2), New Zealand, Norway and Northern

Ireland in the United Kingdom use computer-based uniform technology for the actual administration of central assessments (and for central examinations in the case of Norway).

Denmark is the only country participating in the OECD Review which reported using computer-based adaptive technology for its national assessments in Danish (reading), English, mathematics, biology, geography and physics/chemistry for different year levels. In these adaptive assessments, test items are selected sequentially according to a student's performance on the previous test items. This makes testing more efficient as more fine-grained information can be obtained in less testing time. As reported in the Dutch Country Background Report, the Cito pupil monitoring system in the Netherlands also contains computer-based tests some of which are adaptive (Scheerens et al., 2012).

In Australia, through the governments' Online Diagnostic Tools initiative, the potential to deliver national assessment programmes on line is also being explored through detailed research and consultations on the measurement and assessment effects of online testing and through discussions with school authorities regarding technical and delivery considerations. The delivery of national online assessment offers the opportunity to further integrate ICT into teaching and learning and provide more individualised diagnostic assessment of student progress.

Research and development in ICT-based assessment

While the systematic use of ICT to transform central assessment systems is still limited, many public and private actors are increasingly investing in research and development in this area. Several innovations in assessment around the world are taking advantage of recent advances in ICT. Increasingly sophisticated ICT programmes that score "open-ended performances", such as essays, are under development (see Chung and Baker, 2003; Chung et al., 2001; Herl et al., 1999, in Looney, 2009). These programmes use natural-language processing, artificial intelligence and/or information retrieval technologies to detect textual features of essays (for example, variety in use of syntax, quality of content and organisation of ideas). These ICT models are still in the relatively early stages of development however, and while they may facilitate scoring of large-scale assessments, cannot replace human raters. Further studies are also needed to determine the validity and reliability of different automated essay scoring tools (Wang and Brown, 2007, in Looney, 2009).

Technology-based assessments may also incorporate simulation, interactivity and constructed response formats. For example, students may use the multimedia functions of ICT to show how they would perform a science experiment or other problem solving tasks (e.g. the Virtual Performance Assessment project at Harvard University, Box 4.10). They may use Internet–based programmes to "predict-observe-explain" specific concepts. Or, they may develop concept maps using online tools to show their understanding of processes. The student's map can then be scored by comparing it against an expert map (Bennett, 2001, in Looney, 2009). There are also examples of web-based peer assessment strategies (Binkley et al., 2010). Research is also undertaken in the area of assessments that can measure collaborative problem-solving. For example, students can send information and documents to each other and work on tasks together using ICT (Binkley et al., 2010).

In another example of ICT-based assessment, students are scored on their use and judgement of information on the Internet. Students may obtain feedback on their work in real time. Examples of tests that track student activities on a computer while answering a

question or performing a task are the iSkills test developed by the Educational Testing Service (ETS) in the United States, and the ICT literacy test which is part of Australia's National Assessment Program (Box 4.10). Test developers have found that students respond positively to Internet tasks (searching, and judging the quality of on-line information) and are engaged, even in difficult, open-ended tasks (Lewis, 1998, in Looney, 2009).

Box 4.10 Sophisticated ICT-based assessments

The **Virtual Performance Assessment (VPA)** project at Harvard University (*http://vpa.gse.harvard.edu/*) in the United States uses innovations in technology and assessment to measure students' ability to perform scientific inquiry to solve a problem. The virtual assessments use similar design as videogames to support students' experimentation and problem-solving skills. Participants take on the identity of a virtual scientist and can walk around the environment, make observations, gather data, and solve a scientific problem within a context. The student's arguments are expressed through various digital media including concept maps, data collection tools and conclusion tools for submitting hypotheses and causal statements. The student's actions are stored in logfiles and can be used to track students' inquiry trajectories.

The **iSkills test** developed by the Educational Testing Service (*www.ets.org*) in the United States intends to measure constructs not directly accessible through conventional means, testing the ability to use technology as a tool for cognitive purposes. The iSkills test intends to measure students' critical thinking and problem-solving skills in a digital environment. In a one-hour exam real-time, scenario-based tasks are presented that measure an individual's ability to navigate, critically evaluate and understand the wealth of information available through digital technology. The programme provides individual and group data for use in student evaluation and placement.

The **National Assessment Program** in Australia includes an assessment of students' ICT literacy. It is designed as an authentic performance assessment. The assessment is intended to mirror students' typical "real world" use of ICT. In the 2005 and 2008 rounds of the assessment, students completed tasks on computers using software that included a seamless combination of simulated and live applications. Some tasks were automatically scored and others (those that resulted in information products) were stored and marked by human assessors. The tasks (items) were grouped in thematically linked modules, each of which followed a narrative sequence covering a range of school-based and out-of-school based themes. Each module typically involved students collecting and appraising information as well as synthesising and reframing the information. The assessment involved a number of modules so as to ensure that the assessment instrument assessed what was common to the ICT literacy construct across a sufficient breadth of contexts (MCEECTYA, 2008). The 2011 ICT Literacy report is now also available.

The New Zealand Ministry of Education (2011) has developed digital portfolio guidelines "for beginners". These guidelines aim to provide information on e-portfolios to non-technical users and can help school leaders consider the place of e-portfolios in their school's educational strategy. The guidelines also provide an overview of available tools and case studies of schools having implemented the use of e-portfolios. The Ministry of Education has also supported the development of the e-portfolio service **My Portfolio (Mahara)** (*http://myportfolio.school.nz/*), which provides a personal learning environment to record and showcase evidence of achievement, manage development plans, set goals, and create online learning communities. New Zealand Schools can register free of charge. Most New Zealand universities have introduced MyPortfolio to their teacher training programmes and some are now familiarising their schools' advisers with MyPortfolio. The Ministry of Education is also offering "taster" sessions to groups of teachers with an interest in using MyPortfolio.

Sources: Binkley et al. (2010); European Commission (2011a); MCEECTYA (2008), New Zealand Ministry of Education (2011).

Recent developments in ICT are relevant not only for standardised assessments, but they can also influence regular assessment practices in the classroom. For example, there has been increased interest in using digital portfolios (or e-portfolios) across countries (McFarlane, 2003; Binkley et al., 2010; Pepper, 2011). While portfolios have been used in many countries for some time, the use of digital tools allows collecting information on student progress in a broader range of formats including text with hyperlinks, video, audio and simulations. Digital portfolios also make it easier for teachers to comment on assignments and track student progress. Students' own work with the digital portfolios can enhance their skills in learning to learn, ICT literacy and self-monitoring (Binkley et al., 2010; Pepper, 2011). The New Zealand Ministry for Education, for example, developed e-portfolio guidelines for non-technical users and supported the development of an e-portfolio service called MyPortfolio (Mahara).

ICT-based assessment also provides opportunities for more equitable and adapted assessment of diverse students. The degree of personalisation possible in ICT-based assessment is greater than in traditional assessment given the range of presentations, response formats and contexts available, for example through item banks, test practices and e-portfolios. The Computers for Pupils initiative developed by the UK Department for Children Schools and Families (now Department for Education), constitutes one attempt to increase equity through e-Assessment. Through providing e-learning and e-assessment in the homes of disadvantaged children, this initiative has increased their participation in assessment (e.g. through handheld and haptic technologies). Furthermore, this initiative has provided students with a platform for social networking and self-help groups. In such communities, students may get in contact with peers that experience the same challenges and provide each other with mutual support (Nayral de Puybusque, forthcoming).

Equity and fairness in assessment design

It is important that assessments allow all students to show what they know and can do without being unfairly hampered by individual characteristics that are irrelevant to what is being assessed (Binkley et al., 2010). Assessment needs to be appropriate for students at the range of developmental levels likely to be in assessed population and sensitive to the needs of particular groups such as cultural minorities, students whose mother tongue is not the language of instruction (L2 students) and students with special educational needs. In the process of developing student assessment, notably standardised testing, it is often unclear whether accessibility for specific student groups such as cultural minorities or students with disabilities has received sufficient attention.

The research literature on bias and equity in assessment is extensive, but it often focuses on specific sub-groups, in particular disabled students and L2 students. To ensure fairness in assessment for all students, it is important to develop frameworks for equitable assessment for the wide range of different sub-groups without privileging one group over another. The development of a broad framework for equity in assessment for *all* children requires central guidelines for orientation and coherence across educational settings, but it should at the same time allow for flexibility and adaptability of practices at the local and school level (Nayral de Puybusque, forthcoming).

The design of assessments is open to a number of risks for equity namely if it is biased for particular student groups. Test bias refers to differential validity of a test between specific sub-groups of students (Sattler, 1992). There are several types of potential bias. For example, *irrelevant context* may bias a test against a certain group of students. Unnecessary linguistic complexity is one example for context-irrelevant bias in

assessment, particularly when testing students who do not speak the language of instruction and assessment at home (L2 students). Linguistic complexity slows the student when reading and listening in comprehension and answering, gives room for misinterpretations, and increases the cognitive load (Nayral de Puybusque, forthcoming).

There may also be *bias in content validity* when the choice of a particular set of knowledge and skills is likely to privilege certain groups of students over others (Klenowski, 2009). In this case, after the general ability level of the two groups is held constant, the test is relatively more difficult for one student group than another (Reynolds, 1998, in Whiting and Ford, n.d.). For example, if asked a question about hockey, a student or group who has never played or watched or had discussions about hockey is at a disadvantage. The lack of exposure and experience in relation to particular content places them at a disadvantage (Whiting and Ford, n.d.). There may also be *bias in item selection*, which is related to how one item is included in the test while another is not. While an overall test may not be biased statistically, a few items in it may be. Finally, the choice of method may also lead to bias for certain groups, depending on their familiarity with the general idea of a test, the motivational context in which the test is taken and the frequently implicit assumptions about appropriate behaviour in such a context. For example, students who are more familiar with multiple-choice tests may have developed better strategies to deal with this assessment method than students who have never been confronted with this format before.

Reviewing equity dimensions in test design

While reliability and validity are necessary conditions for any effective assessment system, one cannot assume that these conditions are met or transferable to all different subgroups of the population. Evidence of differential validity is required to determine whether separate test validities are needed for each group (Shultz and Whitney, 2005).

"Equity scanning" is a statistical approach to address testing bias. This method provides the information and means for test developers to identify and eliminate inequitable elements in a test and ensure its quality and the integrity. Differential Test Functioning (DTF), Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) constitute techniques that evaluate a test when applied to a heterogeneous group of students. All of these three methods aim to assess the technical quality, i.e. the validity of a test. The different methods look at variation of performance across sub-groups on different levels of bias. DTF examines a test as a whole; DIF analyses a test at the item level; and DDF focuses on distractors (i.e. incorrect options in a multiple-choice test). Where two students with the same knowledge, abilities and skills turn out to perform differently at a test in terms of scores, answer, or preferred choice, these techniques may be applied. However, the effects and benefits of these methods have been debated and equity scanning still constitutes mainly an area of academic enquiry rather than a field of policy implementation (Nayral de Puybusque, forthcoming). Statistical methods are often only applied ex-post and, therefore, do not address the problem of inequity in assessment directly. Some authors have also criticised that these measures are more likely to detect test anomalies rather than detecting real bias (Battiste, in Volante, 2008).

"Judgemental reviews" constitute another technique to eliminate culturally biased items in a test. Teachers from diverse backgrounds may be allowed to screen a test for cultural bias within a review process. However, it is important to bear in mind the significant impact of the respective teacher's subjectivity when evaluating this method. To address both construct and method bias, it has also been proposed to use informants

with a thorough knowledge of the respective culture and language of a sub-group (van de Vijver and Tanzer, 2004). In the "committee approach", a group of people with different areas of expertise (such as cultural, linguistic, and psychological) review a test or a test translation. The major strength of the committee approach is the co-operative effort among members with complimentary areas of expertise. In Australia, consultations were conducted regarding the accessibility of the National Assessment Program – Literacy and Numeracy (NAPLAN) for Indigenous students. The NAPLAN test development process involves the consultation of Indigenous experts in states and territories. They provide specific feedback on the suitability of the test items and the appropriateness of the stimulus materials for Indigenous students. Test trials are also carried out using a sample of students, and analysis of the results is undertaken to ensure that all items are culturally appropriate and free of bias (Santiago et al., 2012).

Providing accommodations to specific student groups

The tension between standardised tests that are supposed to be common across all students and the need to be sensitive to local cultural knowledge and individual student conditions is a difficult one. One option to provide fair assessment, while upholding common standards, is to provide accommodations for specific student groups. Accommodations aim to "remove causes of irrelevant variance in each student's test performance, thereby producing a measure of each student's knowledge and skills that is valid" (Thurlow et al., 2008). Such measures are not expected to decrease learning expectations for these students but rather to reduce or even eliminate the effects of a student's disability or disadvantage in relation to a particular assessment (Thompson et al., 2005).

Hopper (2001) described accommodation through four aspects: presentation, response, setting, and scheduling/timing. *Presentation* refers to alternate modes of access to the information adapted to the student's particular condition (e.g. a disability or a limited proficiency in the assessment language). *Response* accommodation allows for alternate ways of completing an assessment, with the possibility of using other materials or devices. *Setting* accommodation allows for different assessment conditions in terms of assessment location and environment. *Timing and scheduling* accommodation allows for taking a test at a different point in time, for a different length of allocated time (i.e. including extra-time), or its organisation (e.g. reallocation of time lengths depending on the task). One specific method of accommodation can address one or also several of these dimensions. The use of ICT-based assessments has great potential to provide a range of accommodations for specific student groups.

For sensory disabilities (i.e. hearing and visual impairments), the most widely used accommodations lie in the field of presentation and response. Examples for the various measures used include Braille or large-print papers, a reader, audio playback, a writer, audio recording of oral responses, headphones, a sign-language interpreter, or additional time. In the case of physical and cognitive disabilities electronic devices are common, but setting accommodations are most widely used. For students with physical disabilities (including reduced mobility), access to different test centres, to appropriate seating for a student to be able to take medication and computer devices constitute typical forms of accommodation. For students with behavioural disabilities, supervised rest breaks, permission to move around, and additional time, especially for dyslexic students, are widely used. For students with dyslexia or learning disabilities, spellchecker software can also be available (Nayral de Puybusque, forthcoming).

It should be noted that evidence on the benefits of using different types of accommodations in assessment is often lacking or not conclusive. Abedi et al. (2004), for instance, raise caution regarding the use of dictionaries in assessment. Depending on the brand, the content and the vocabulary level of the definitions, dictionaries may be more or less useful to a student. L2 students must be able to understand the definitions and be familiar with the use of a dictionary. For assessment accommodations to be most helpful to students, it is important that they are well aligned with instructional accommodations. To this end, some education systems require proof that the demanded accommodations have been provided and used in class and/or in internal examinations (Thurlow et al., 2008).

Assessment accommodations for second language learners

Accommodation with regards to the language of assessment is a crucial, but at the same time very complex concern. For important language minorities, several countries have attempted to provide a range of options for students to be assessed in their first language. This can be done by administrating the assessment orally involving an assessor who is familiar with the student's linguistic background. Another possibility is to translate or develop assessment instruments in the student's first language. There are several options for doing this: One would consist in translating and adapting most existing tests which are publicly available. Another option would be to develop instruments in the minority language. Conceptual equivalence would be accomplished by having bilingual teachers and experts working together to moderate the test construction in each language. Assessment instruments developed in this context would have no metric equivalence and intergroup comparisons would not be possible. The main benefit of this procedure, however, would be to reduce the risk of any cultural or linguistic bias.[8] A third option would be to find an intermediate solution between the two previous ones, by developing anchor points in assessment instruments developed in two different languages.[9]

However, assessment in a student's first language, whether through oral administration or translation, may entail several limitations and requirements when students are not actually taught in that language. Several authors have argued that first language assessments are useful only if students can demonstrate their performance more effectively in their first language, typically because they have received relevant instruction in that language (Abedi et al., 2004). To achieve alignment between the language of instruction and the language of assessment, a test needs to reflect the terminology and language used in content-area instruction (Abedi, 2004).

Reducing linguistic complexity is another frequently used approach to eliminate bias against L2 students in assessment. Low-frequency vocabulary, passive voice verb constructions and the length of written constructions feature among the main challenges for L2 students (Abedi and Gandara, 2006). Accommodation measures to reduce the linguistic complexity of a test include simplification, modification and translation of test elements (e.g. items, directions) as well as the use of dictionaries, and glossaries. Linguistic modifications of content-based items without changes in the content-related terminology represent one possible intervention.

At the same time, it is important to bear in mind that the state of the empirical evidence on the validity, effectiveness and efficiency of many accommodation measures varies and is at times contradictory. In the various states of the United States, Rivera et al. (2006) identified 75 different accommodation tools allowed for L2 students (Rivera et al., 2006). Out of these, only 44 were recognised to fit student's needs. Abedi (2004) judged only 11 measures as effective and appropriate. Another field where progress needs to be made

relates to the diversity of the L2 sub-group and equity within this group. In most countries, it is not cost-effective to develop translated assessments in all the many languages spoken by learners at home, which introduces questions of fairness and equity for smaller language groups. In the United States, for example, while the large presence of the Spanish-speaking minority has brought benefits for L2 students as a whole, equity towards smaller L2 groups is an important issue that should not be neglected (Abedi et al., 2004).

Improving equity through multiple assessment opportunities

Several studies indicate that certain formats of assessment may advantage or disadvantage certain student groups (Gipps and Stobart, 2004). Since it is very difficult to make assessment wholly inclusive and neutral – any format, content, construct, and method may be biased in some direction – a mix of different versions of format, content and construct may help ensure fairer assessment. High-stakes decisions about students should not be based on the results of one test alone. An important approach to offering fairness is to collect multiple data, use a range of assessment tasks involving a variety of contexts, response formats and styles and draw on this comprehensive approach to make decisions. This broader approach is likely to offer students alternative opportunities to demonstrate their performance if they are disadvantaged by any one particular assessment in the programme (Gipps and Stobart, 2009).

It is also important that the format and design of different assessment instruments is informed by research on effective approaches for diverse student groups. The Australasian Curriculum, Assessment and Certification Authorities, for instance, recommended the distribution of "summaries of relevant research findings to item writers and test designers" (ACACA, 1995). In areas where there is limited research, for example inclusive assessment for students from cultural minorities, it is important for education systems to encourage the development of more research to extend the knowledge and evidence base.

Finally, equity in assessment is also about creating new attitudes, mentalities, and skills at every level of the educational system. This is a long-term process. As will be discussed below in the section on "Capacity", developing the competencies of teachers for inclusive assessment is key to avoiding bias in teacher-based assessment and improving equity in assessment.

Capacity

This third section discusses the competencies necessary to assess students, to benefit from assessment and to use the results of student assessment. It includes issues such as: the capacities students need to engage in and benefit from their assessment; the assessment competencies that teachers acquire in initial teacher education, professional development and moderation arrangements; and the expertise of the agencies involved in student assessment.

Student capacity for assessment

Traditionally, teachers have been regarded as responsible for establishing where learners are in their learning, where they are going, and what needs to be done to get them there. In recent years, there has been increasing focus on the role of the learner in assessment, not only as an active participant, but also as the critical connector between assessment and learning (Butler and Winne, 1995, in McDonald and Boud, 2003; Earl, 2003). While feedback by teachers and others provides information that can help students

improve, it is the students themselves who must make sense of that information, relate it to prior knowledge and take action to close gaps in their own learning. This is the regulatory process in metacognition. It occurs when students personally monitor what they are learning and use this monitoring to make adjustments in how they learn and understand.

Self- and peer-assessment are powerful processes that aim to enhance the role of learners in their own assessment. Self-assessment has been defined as "the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they met these criteria and standards" (Boud, 1986). Peer-assessment, on the other hand, involves students in assessing each others' work, again through reflection on goals and on what it means to achieve them. It may take place in pairs or in groups and has particular value in formative assessment as it allows students to use familiar language and ask each other questions that they may not dare to ask their teachers (EPPI, 2005). Such approaches to assessment can promote a greater sense of agency and responsibility of students in their own learning, and can help them engage in fruitful conversations about their learning leading to greater self-confidence, metacognitive monitoring skills and self-regulation skills, sometimes referred to as "assessment *as* learning" (Earl, 2003). A range of studies reviewed by Black and Wiliam (1998) report positive effects of student self-monitoring on the learning of different student groups (Sawyer et al., 1992; McCurdy and Shapiro, 1992; Masqud and Pillai, 1991; Merret and Merret, 1992).

Little information is available internationally regarding the extent to which students are engaged in their own assessment across countries. Some countries are beginning to implement policy frameworks that emphasise the importance of building the learners' own capacity for self-assessment and self-monitoring. In Ireland for example, the National Council for Curriculum and Assessment (NCCA) guidelines on assessment emphasise the importance of sharing learning goals with learners, helping learners to recognise the standards they are aiming for, involving learners in assessing their own learning, providing learners with feedback, communicating confidence to learners that every learner can improve, and adjusting teaching to take account of the results of assessment.

In Canada, several jurisdictions highlight the role of the learner in assessment within their curricula. In the Manitoba English Language Arts Curriculum for Years 5 to 8, for example, it is stated that:

> Modelling and encouraging metacognitive strategies helps students to understand, monitor, and direct their learning processes. Metacognitive questions such as, "What do you notice about your thinking?" and "How did you remember that information?" help students develop internal conversations and reflection about the learning process. When students have opportunities to reflect on their learning, especially with peers, they begin to develop self-assessment skills and want to take more responsibility for shaping and directing their own learning experiences. At times, students need quiet reflection. Whether alone or with others, students use reflection as a tool to consolidate what, how, and why they have learned, and to set goals for future learning.
>
> *(*Manitoba Education website, English Language Arts, Curriculum Documents, *www.edu.gov.mb.ca/k12/cur/ela/curdoc.html.)*

However, in many countries participating in the OECD Review, there were concerns that formative assessment approaches tended to remain teacher-centred rather than student-centred. While self- and peer-assessment are beginning to receive increasing attention in both policy and practice across countries, it appears that in many contexts

self-assessment is understood in a context of self-marking rather than reflection about learning. In several countries where OECD Country Reviews were organised, self- and peer- assessment practices were incipient, with little attention to sharing and co-constructing learning goals and criteria with students and involving students in reflecting on their progress and evaluating their learning outcomes.

To help students monitor their own learning, it is essential that they understand the learning goals and what they need to do in order to reach them (Sadler, 1989). Assessment schemes and purposes, as well as the specification of what will be assessed and against which criteria the judgement will be made, must be transparent to students (Ross et al., 1999). As students internalise the criteria for evaluating their work, they are better able to connect their performance with their preparation, and develop an internally oriented sense of self-efficacy (Stiggins, 2005). Teachers can use classroom assessment as the vehicle for helping students develop, practice, and become comfortable with reflection and with critical analysis of their own learning (Earl and Katz, 2008).

Teachers themselves also need to learn how to develop learner-centred teaching and assessment and how to introduce self-assessment practices into regular classroom activities. In a study conducted in Barbados, McDonald and Boud (2003) found positive effects of formal self-assessment training for teachers and students on student performance in external examinations. The study was conducted in the context of a large-scale introduction of self-assessment across a range of subjects, where teachers were trained in self-assessment practices and introduced these to a group of students preparing external examinations. The performance of students participating in this process was compared with that of a matched control group of students who were not given self-assessment training. The authors found a significant difference in performance with those trained in self-assessment outperforming the control group in each curriculum area.

Teacher capacity for assessment

How to best prepare teachers for their assessment responsibilities is the subject of debate in many countries because of the complexity of assessment and its integrated role with understanding teaching and learning. In the OECD's Teaching and Learning International Survey (TALIS), across the 23 participating countries, 15.7% of teachers indicated having "high professional development needs" in the area of student assessment practices in 2008. However, considerable differences can be observed across countries, with larger proportions of teachers (above 20%) in Brazil, Korea, Norway, Slovenia, Italy, Lithuania and Malaysia expressing high needs in this area (OECD, 2009).

Little comparable information is available internationally regarding the preparation and training that teachers receive to build their assessment capacities. In most countries, institutions responsible for teacher education are autonomous and define their own curricula, which naturally leads to variations across institutions regarding the content of initial teacher education. Information from education systems participating in the OECD Review provides some indications about the place given to assessment approaches within initial teacher education and professional development opportunities across countries.

Initial teacher education

Broadly, the information collected from education systems through the OECD Review appears to show that student assessment is given increasing attention in initial teacher education programmes internationally. Initial teacher education programmes vary

in the way that assessment is taught: *(i)* in a dedicated assessment course; *(ii)* within curriculum areas; *(iii)* theoretically; and *(iv)* practically.

In a few countries, such as Mexico, initial teacher education is centrally organised with a nationally uniform curriculum for teacher education university degrees. In Mexico, according to the national curricula for basic and lower secondary education, teachers should be trained to perform student assessment in the classroom during their initial education. Capacity to assess student learning is supposed to be developed through a course called "teaching planning and learning assessment", taught six hours per week in the sixth semester of the Bachelor's in Primary Education and four hours per week in the fourth semester of the Bachelor's in Lower Secondary Education. This course should include both assessment of learning processes and formative assessment practices (Santiago et al., 2012c).

Several countries report that while initial teacher education institutions are autonomous, they typically provide prospective teachers with basic knowledge and skills in student assessment approaches. This is the case for example in Australia and New Zealand (Nusche et al., 2012; Santiago et al., 2011). In Canada, many universities offer courses at the undergraduate level dealing with assessment practices. Some are compulsory courses for students working on becoming certified teachers, while others are offered on line for further professional development (Fournier and Mildon, forthcoming). In Finland, the assessment of students' progress and learning is taught in initial teacher education, as well as in vocational teacher education (Finnish Ministry of Education and Culture, forthcoming). In Korea, subjects such as educational evaluation, measurement and assessment of education, education research methodology, psychological examination, educational statistics and psychological measurement are provided as compulsory or optional courses for prospective teachers. Educational contents include the basic concept of educational evaluation, classification and types of assessment, principles and practice of test development, principles and planning of performance evaluation, sufficiency rating for test items, basic statistical analysis, results utilisation, and general classroom and student assessment (Lee et al., 2004, in Kim et al., 2010).

In a range of education systems, there have been central decrees or initiatives to restructure or regroup teacher education institutions, which allowed the central authorities to emphasise particular priority topics including student assessment. In the Flemish Community of Belgium, the Institutes for Initial Teacher Education were restructured by decree in 2006. According to the decree, the Institutes' curricula need to guide prospective teachers towards basic competencies, including skills for student assessment (Flemish Ministry of Education and Training, 2010). In the French Community of Belgium, initial teacher education has been redefined by decree in 2000 and now includes 30 units dedicated to assessment, differentiation of learning, identification of learning difficulties and definition of remedial strategies. Assessment approaches are also part of the pedagogical competencies to be covered throughout the initial teacher education for secondary teachers (Blondin and Giot, 2011). In Ireland, the content and duration of initial teacher education courses have been reconfigured as part of measures introduced under the National Literacy and Numeracy Strategy (Irish Department of Education and Skills, 2011) and the enactment of the Teaching Council Act 2001. The Teaching Council has also published criteria and guidelines which providers of initial teacher education are required to observe. In Norway, a framework plan for a new initial teacher education launched in 2010 provides guidelines regarding the development of teachers' assessment competencies. In particular, it requires that assessment for learning should be one of the competences that teachers have acquired upon graduation. In Sweden, a new initial

teacher education approach was to be implemented from July 2011. The new programmes contain specific goals related to assessment and marking. Assessment topics are also expected to be integrated into the didactics of every subject (Nusche et al., 2011b).

In Norway, the Directorate for Education and Training has also stimulated co-operation among teacher education institutions regarding the topic of student assessment. Until 2011, it funded the Norwegian Network for Student and Apprentice Assessment (NELVU), a network of teacher education institutions that aims to build capacity regarding student assessment within schools and university colleges. To this end, each teacher education institution has formed assessment experts within the institution to work with faculty on this particular topic. The focus was on all aspects of assessment literacy including the use of national test results, assessment for learning and different classroom assessment approaches. NELVU further aimed to stimulate research and development regarding assessment and co-operated with experts internationally, such as the Assessment Reform Group in England (Nusche et al., 2011a).

A recurrent concern across education systems that received an OECD Country Review was that there were variations in the degree to which different teacher education institutions made student assessment a priority in their programmes. In Australia, a survey of teachers revealed that "methods for assessing student learning and development" were among the areas of greatest need for professional development as identified by teachers (Santiago et al., 2011). In the Flemish Community of Belgium, the Inspectorate found in 2007 that starting teachers experienced that developing adaptive or alternative assessment approaches was one of their most difficult tasks, in spite of previous efforts to restructure teacher education with greater emphasis on such approaches (Flemish Ministry of Education and Training, 2010). In the Czech Republic, according to Santiago et al. (2012b), there is very little attention given to developing such assessment skills in initial teacher education programmes. Teachers are more familiar with using test score information for summative purposes and have not received significant training regarding the use of richer assessment tasks to inform their teaching. In Denmark, the OECD review team interviews revealed that pre-service teacher education programmes offered little training in student assessment for teacher candidates (Shewbridge et al., 2011).

Teacher professional learning

Teachers' professional learning regarding assessment is a career-long experience that needs to be sustained. In parallel to changes in initial teacher education, several countries have introduced a range of professional development and learning opportunities for teachers regarding assessment practice. Timperley (2011) describes the difference between professional development and professional learning. Over time, the term "professional development" has taken on connotations of provision of information to teachers in order to influence their practice whereas "professional learning" implies an internal process in which individuals create professional knowledge through interaction with this information in a way that challenges previous assumptions and creates new meanings.

The organisation of professional development courses regarding assessment is common practice in many countries and it appears essential in supporting a national focus on effective assessment practice. Gilmore (2008) makes a distinction between professional development programmes in which assessment is "foregrounded" (i.e. it is the main focus of the programme) and those programmes where assessment is "backgrounded", (i.e. the programme does not focus on assessment per se, but assessment

is an integral part of the programme). Many countries use a mix of both. In several settings, for example Hungary, Mexico and Norway, initiatives are also directed at school supervisors and/or school owners.

In Australia, most jurisdictions provide training to improve the competency of teachers to analyse and interpret student assessment data. For example, the Victorian Curriculum and Assessment Authority conducts in-service courses in schools around Victoria, to develop school leaders' and teachers' skills in interpreting the results of the national assessments and the Victorian Certificate of Education exam (Santiago et al., 2011). In the Flemish Community of Belgium, courses on assessment are an important part of the wide range of in-service training possibilities. It is common practice for schools to invite experts on various items (e.g. student assessment) to provide training opportunities for teachers (Flemish Ministry of Education and Training, 2010). In the French Community of Belgium, professional development on assessment is also available. The subjects dealt with in in-service training courses are determined based on needs identified by the *Commission de pilotage* and the Inspectorate (Blondin and Giot, 2011). In Korea, in-service training on educational assessment is provided as part of the national training framework. In recent years, local education offices and individual schools have also been adding new dimensions to the contents and methodology of such training (Kim et al., 2010).

In New Zealand, the Ministry of Education has initiated several major professional development programmes, which have been evaluated in terms of their impact on student learning, with promising results (Nusche et al., 2012). For example, "Assess to Learn (AtoL)" is a whole-school professional development programme that has been offered to primary and secondary schools since 2002. Schools can apply for participation in the programme and typically participate for two years. The annual budget for AtoL is NZD 3.17 million annually and currently involves 155 schools. The programme intends to support teachers in choosing adequate assessment tools and analysing assessment information so as to further advance student learning. A 2008 evaluation of the AtoL programme reported a significant impact of the programme on teacher professional practice and important improvements in student learning, especially for students with initially low achievement levels. Monitoring data showed that schools participating in AtoL had achieved up to 4.5 times greater shifts in writing achievements in Years 4 to 9 than the nationally expected rate of progress.

In some countries, the focus on professional development for student assessment has been considerably reinforced in recent years. In Hungary, for example, awareness raising campaigns with professional content on assessment were held in 2009 in every region. These occasions can also serve to strengthen the reputation and acceptance of assessments, and give professional impetus to make the use of results as diverse as possible. In Mexico, assessment-related topics are receiving increasing emphasis in the offerings available to teachers. While two years ago only two programmes were specifically focused on assessment issues, the 2011/12 catalogue includes over 30 programmes, among about 1 100 offerings. Most of them are targeted at school supervisors and focused on competencies-based assessment. Simultaneously, many subject-specific courses include new approaches, techniques and instruments for classroom-based assessment (Santiago et al., 2012c). In Norway, student assessment is also being highlighted as a key topic for the continuing professional development of school professionals and school owners. Since 2005, the Directorate for Education and Training has included student assessment as one of the annual priorities for continuing

professional development of teachers, school leaders and trainers of in-service training providers (Nusche et al., 2011a).

In several countries, professional development also takes place through moderation of teachers' assessment and marking. As explained above, moderation refers to quality assurance measures that seek to increase the consistency of marking, for example through teachers reviewing or cross-marking each other's assessments within a school or across schools or working together in groups to discuss assessment criteria and student performance levels. Moderation is a key strategy in validating consistency of teacher judgement and marking and it may occur within schools, between schools and across school sectors. At the same time, moderation also involves professional discussions between teachers about the quality of authentic pieces of student work and as such it has the potential to provide a powerful professional learning opportunity for teachers that they can relate closely to their classroom practices. It also contributes to improving teachers' professional judgements about student work and their developing a shared understanding of marking criteria or standards within schools and between schools (Timperley et al., 2008). This provides teachers with a chance to reflect on assessment in their subject, both on topics and criteria.

Finally, professional learning may also build on existing initiatives, provide opportunities for teachers and schools to network among each other or with assessment advisors and disseminate effective practice. In Norway, the Better Assessment Practices project (2007-09) supported a range of local projects to improve assessment practice in Norwegian schools. As a follow-up, the Assessment for Learning programme (2010-14) was implemented to support school projects and networks focusing particularly on formative assessment. There are also local initiatives in this area. The City of Oslo, for example, employs two "assessment advisors" that schools can invite to provide help regarding assessment (Nusche et al., 2011a). In Denmark, the availability of resource teachers at schools provides important support to teachers. Although assessment and evaluation advisors are few (reported in only 8% of schools) they have the potential to offer critical support to teachers whose initial training did not give particular emphasis to student assessment and evaluation (Shewbridge et al., 2011). In Canada, all Boards have created assessment divisions or sectors within their administrative structure and have assigned personnel to lead workshops, develop activities related to assessments, and to track, collate, analyse and distribute findings of district-wide assessments to their respective stakeholders (Fournier and Mildon, forthcoming).

However, while professional development opportunities in the area of student assessment appear to exist in most countries, there is often little information available at the central level regarding the extent to which teachers benefit from these offers and regarding the quality of available courses. In education systems where schools and teachers are free to determine the content of their professional development courses, it is often unclear to what extent teachers choose to improve their student assessment methods through such courses.

These concerns are linked to broader challenges in teacher professional development systems. While professional development is receiving increasing policy attention across countries, its provision is often patchy and uneven (OECD, 2005b). Teachers typically receive entitlements and incentives to engage in professional development, but these are not always matched by initiatives on the supply side. In some countries, teachers receive public funding only for courses offered by a few institutions, which can reduce the incentives for innovation and quality improvement (OECD, 2005b). Given the rapidly

changing demands being made of teachers in student assessment, countries may be faced with a lack of qualified trainers and well-designed training programmes in this area.

Guidelines and tools to support effective assessment practice

In addition to classic training courses, there are many other ways for education authorities to promote and support the strengthening of teachers' assessment approaches, for example through the development of guidance and support materials. In Denmark, for example, the electronic Evaluation portal provides a plethora of evaluation and assessment tools for use by teachers, schools and municipalities. This includes, for example, guidance materials on how to assess students in the final examinations in Form 9 prepared by the examiners and subject advisors of the final examinations, including advice and ideas for classroom teaching and criteria for student evaluation and assessment in classroom activities (Shewbridge et al., 2011).

In Ireland, the National Council for Curriculum and Assessment (NCCA) guidelines on assessment provide detailed, practical guidance to teachers on how to use a range of assessment approaches in order to obtain a full picture of a child's achievement. They also provide guidance to schools on the development and implementation of an assessment policy. The NCCA has also contributed to the development of expertise in relation to formative assessment through its curriculum development projects with schools and designed materials that are intended to support teachers and schools in expanding their assessment toolkit for classroom-based assessments (Irish Department of Education and Skills, 2012).

In several countries, guidance materials have been developed with the aim of strengthening teachers' understanding of standards and criteria used in central assessments and examinations. In the Slovak Republic, for example, following the organisation of external assessments, the National Institute for Certified Educational Measurements (NÚCEM) publishes analytical reports for teachers with recommendations for improving the quality of education according to measured findings along with methodological guidelines and collections of tasks. NÚCEM also organises expert seminars for teachers to provide participants with feedback on test results (Hajdúková et al., forthcoming). In Ireland, the availability of marking schemes from State examinations combined with the publication of Chief Examiners' reports supports greater understanding among teachers of expected standards and criteria.

Involving teachers in marking central assessments and examinations

In several countries, professional learning is organised around central assessment or examinations. Hiring teachers in order to correct and mark central assessments may contribute substantially to their understanding of expected standards and criteria. In Canada, teachers are involved in the development of regional, provincial and national tests and the associated materials such as scoring guides, rubrics and the choosing of exemplars for use in scoring. Teachers are involved in all stages of the administration process (including field testing) and they are also trained for the scoring of completed assessments, using the prepared rubrics and exemplars. The result has been the accumulation of experienced test developers who will enhance their own teaching practice and may bring to their school communities a rich source of leadership in assessment.

In New Zealand, teacher professional development related to effective assessment occurs via their strong involvement in scoring student work for the National Education Monitoring Project (NEMP) and for the National Certificate of Educational Achievement

(NCEA). For the National Education Monitoring Project, about 100 teachers are freed from their teaching responsibilities each year to conduct the assessments. They receive one week of training and then administer the tasks over a period of five weeks. The intention is to ground the assessment practice in sound teaching practice and to build and strengthen teacher assessment capacity.

In Norway, professional development also takes place around teachers' marking of central examinations and in moderated marking of oral examinations. In addition, some school owners further support moderated marking processes. In 2010, the municipality of Oslo launched a pilot study in lower secondary education, where they invited all schools to implement a mock exam. The municipality invited 60 teachers from 35 schools to come together to mark the examinations in a moderated marking process in collaboration with expert teachers (who had been part of the national marking process). This provided an opportunity for teachers from Oslo schools to engage in discussion about the meaning of marking criteria in relation to examples of student work.

Concerns about fragmented approaches to teacher learning in assessment

In spite of remarkable efforts across countries to ensure that teachers acquire basic assessment literacy through initial teacher education programmes and continuing professional learning, countries also report that more needs to be done to help teachers develop the competencies necessary for effective assessment. In all countries, there are indications of some inadequacies in teachers' preparation for student assessment. New teachers sometimes need substantial support to develop classroom assessment techniques, reporting schemes and moderation processes, which in some cases are not priority areas during teacher education programmes. More experienced teachers may lack access to continuing professional development supporting them in the use of most recent assessment technology and staying up to date on emerging research on effective assessment approaches. Experience from across OECD countries reveals that the following areas require particular attention:

- **Assessing key competencies**. While internal assessment provides opportunities for diverse and innovative assessment approaches, its validity depends to a large extent on the assessment opportunities provided by individual teachers (Harlen, 2007). Indeed, reviews of research on teacher-based assessment note that teacher-made assessments often focus on content knowledge and may be no more diverse or innovative than external tests (e.g. Crooks, 1988; Black, 1993; Black and Wiliam, 1998; Harlen, 2007). Hence, it is important to ensure that teachers receive adequate training to assess a broader range of skills and competencies (Somerset, 1996).
- **Effective marking and reporting**. There are sometimes large variations in the ways teachers assess students and set marks. This is compounded in situations where there are no central marking criteria and where there is no guarantee that teachers engage in discussion or moderation within or across schools. In the case of summative assessments that carry high stakes for students, this poses important challenges to the fairness of assessment and marking. It is important to provide focussed training on how to make summative judgements on student performance in relation to central curriculum goals or standards. Such training should include how to identify valid evidence and how to apply scoring rubrics and marking criteria to very different types of evidence of student learning (Harlen, 2004).

- **Effective formative assessment**. Ongoing attention to teacher training in formative assessment is also vital. An important priority is to develop teachers' capacity to interpret student assessment data, including that generated by standardised tests, for the improvement of classroom instruction. To become assessment literate, teachers need to be aware of the different factors that may influence the validity and reliability of results and develop capacity to make sense of data, identify appropriate actions and track progress. Other key areas of training in formative assessment are to help teachers provide effective feedback to students and to fully engage students in their own assessment (Looney, 2011a; Earl and Fullan, 2003).
- **Inclusive and fair assessment of diverse students**. Several studies report that there are risks of conscious or unconscious bias in teacher-based assessment (Crooks, 2004; Harlen, 2007). Bias in teachers' assessment may be related to teachers' prior knowledge of student characteristics such as behaviour, gender, special educational needs, immigrant background, first language, overall academic achievement or verbal ability (Harlen, 2004; Gipps and Stobart, 2004). The key challenge is to ensure that rich assessment opportunities are systematically offered to all students regardless of such characteristics. It is important that dimensions of inclusive and equitable assessment are covered in both initial education and professional development to help teachers adapt their assessment approaches to the diversity of student needs in their classrooms.

What is missing in many countries is a *strategic* approach to teacher learning in assessment. While teachers may learn bits and pieces about effective assessment at various stages of their career, the offer of professional learning in this area appears fragmented and limited in scope. The vision of assessment communicated in professional development courses is not always well aligned to the way assessment is covered in initial teacher education. Also, the different professional development programmes regarding assessment are typically run independently of each other, often without recognition of overlaps and synergies. The development of teacher standards or professional profiles – and the inclusion of assessment as an important teacher competency in these standards – can help provide purpose and structure for professional development at different stages of the teaching career (Chapter 5). There is a role to play for the central authorities to ensure alignment of publicly funded professional development courses so that they foster a coherent understanding of the interrelations between curriculum, teaching, learning and assessment (Absolum et al., 2009).

Central capacity for assessment development and implementation

In all education systems, the educational administration and/or a range of central agencies are involved in student assessment in various ways. As explained in Chapter 3, several countries have created specialised assessment agencies in recent years, reflecting the growing importance of educational measurement and the increasing complexity of evaluation and assessment frameworks across countries. In Austria, for example, the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) was created in 2008 to develop, implement and monitor education standards. Similarly, in the Slovak Republic, the National Institute of Certified Measurement (NÚCEM) was created by the Slovak Ministry of Education in 2008 to develop, administer and oversee all national tests and assessments. Other countries have more long-standing agencies in charge of overseeing and developing the education system's assessment strategy. Expertise regarding education standards and assessment is

typically also concentrated in specific departments of the Ministries of education, Inspectorates and education review bodies, curriculum organisations, quality assurance agencies and examinations and qualifications authorities (for a detailed overview, see Chapter 3).

Student assessment is a highly technical matter and the design and implementation of standardised assessments requires expert capacity which takes time to be developed. In several education systems where OECD Country Reviews took place, there were concerns about the lack of expertise at the central level in the area of student assessment. Such expertise is particularly important when education systems are developing and introducing large-scale assessments. Considerable investment is needed to develop capacity and expertise in standardised test development, including areas such as educational measurement and assessment design.

In addition, a deep understanding about the role of assessment in the improvement of policy and practice is essential to develop policy, support the development of central assessment tools and ensure that assessment results are used appropriately for monitoring and support of schools. Central agencies can build such capacity by engaging with each other and with the assessment community more widely, both nationally and internationally. For example, the establishment of independent working groups from a range of sectors and organisations in education can help accompany the development of central assessment strategies, by debating the central assessment instruments, monitoring their implementation and conducting impact evaluations. Such a group should have the remit and expertise to investigate and make recommendations that centrally mandated assessments are valid and reliable.

Some tools are also available internationally to help governments and central agencies review and further develop student assessment systems. The World Bank, for example, has developed a set of tools called SABER (Systems Approach for Better Education Results, available at *www.worldbank.org/education/saber*) to support countries in systematically reviewing the strengths and weaknesses of their student assessment systems. These tools comprise a set of standardised questionnaires and rubrics for collecting and evaluating data on three types of assessment (classroom assessment, examinations and large-scale system-level assessment). Policy makers and central agencies can use these tools to benchmark their system-level policies and practices and plan for further development. However, it is important to note that additional tools will be necessary to determine actual assessment practices implemented in classrooms across a given country (Clarke, 2012).

It is also the role of the central administration and agencies to establish and share a vision for assessment in the school system. This requires the development of a comprehensive communication strategy to engage stakeholders at all levels in dialogue regarding effective assessment. Leadership and communication capacity required at the central level in order to build a common understanding of the different purposes and uses of assessment in education. In Norway, for example, the Directorate for Education and Training has created a department on internal governance to enhance continuous reflection about the uptake and impact of new initiatives for quality improvement in the education sector. It has also introduced a professional development programme to build leadership among its staff and to enhance effective goal-setting and strategy development within the Directorate itself (Nusche et al., 2011a).

Reporting and use of results

This section is concerned with how assessment results are reported and used for both summative and formative purposes. It describes standards of quality and reporting formats used in different contexts, reviews the legal frameworks in place to regulate reporting of results and discusses the ways in which assessment results are used in different contexts to record information, provide feedback to students and make decisions about their further educational trajectory. The ways in which assessment results are recorded and utilised are essential to the distinction between summative and formative assessment.

Regular summative reporting helps communicate summary statements of achievement to students and parents. As such, it can help reach out to parents and involve them in supporting their children's learning. Records of student achievement can also help teachers communicate about student achievement within a school and make decisions about targeted support. They can contribute to ease transitions when students are changing schools and to ensure consistency between different levels of education. In many countries, such records are also used to make decisions regarding the student's educational trajectory, for example for decisions regarding school entry, year repetition, transfer and ability grouping.

The formative collection and use of results, on the other hand, is embedded in the process of teaching and learning itself. The distinguishing feature of formative assessment in all countries is that the information is used to make improvements (Bloom, 1968; Scriven, 1967). But the way in which information is used and the timescale for decisions may vary across teachers, schools and education systems. Wiliam (2006) distinguishes between long-, medium, and short-cycle formative assessment. According to Wiliam, long-cycle formative assessment occurs across marking periods, semesters or even years (four weeks to a year or more); medium-cycle formative assessment occurs within and between teaching units (three days to four weeks); and a short-cycle formative assessment occurs within and between lessons (five seconds to two days).

Overall, the utility and sound use of data, of course, depends on teachers' assessment literacy and ability to appropriately integrate assessment data and learning in classroom instruction, including the appropriate use of standardised tests. This means that teachers and school leaders need to continually develop their capacity to collect and report on student assessment to students, parents and external partners in effective ways without oversimplifying the complex issues involved in student learning.

Standards of quality and reporting formats

The interpretation of any kind of measurement depends on the standards of quality that are applied. A student's performance may be measured in three different ways:

- **Criterion-referenced** (performance in relation to established standards or criteria). Criterion-referenced assessments are used to make judgements about absolute levels of performance. Such assessments may set benchmarks for what constitutes "mastery" or "high performance" and/or determine minimum standards that should be achieved by every student.
- **Norm-referenced** (performance in relation to a defined group). Norm-referenced assessments classify students based on a comparison among them. The results of norm-referenced assessments have meaning only in comparison with the results of

other students. They do not reflect their proficiency in relation to absolute standards, but in relative terms.

- **Self-referenced** or **ipsative** (change in performance over time). Self-referenced assessments are generally used formatively by teachers to track the growth and progress of individual students over time.

Countries tend to use a mix of different quality standards and reporting formats depending on the specific purpose of each assessment. It can be argued that criterion-referenced assessments are more useful to inform future teaching and learning because they measure student performance against specific learning goals. Norm-referenced assessments, on the other hand, may be more useful for the purpose of student selection (e.g. for university admission) because they allow to compare students among each other (Looney, 2011a). Criterion-referenced assessments tend to be more common for the purpose of summative examinations in lower secondary education (OECD, 2011). Of 13 countries for which information was available, only 2 were using norm-referenced assessments, while 10 were using criterion-referenced assessments. In the United States, both criterion-referenced and norm-referenced assessments were allowed with decisions taken at the school level. A similar picture can be observed in upper secondary education: of 19 OECD countries for which information was available, only 5 were using norm-referenced central examinations while 14 countries were using criterion-referenced examinations (OECD, 2011).

For assessment systems to be well aligned with curriculum goals, teachers are generally expected to assess students in relation to central curriculum goals and standards (Table 4.1). However, the OECD Country Reviews indicate that where assessment standards and criteria are not clearly formulated for different subjects and year levels, teachers often use their own personal reference points, based on their experience and school-based expectations. Teachers' classroom-based assessments often are a mixture of norm-referenced (in relation to other students), content-referenced (in relation to what they taught) and self-referenced (progress of individual students). In some countries, there were concerns that teachers tended to give norm-referenced marks by comparing a student with other students in a class. Such comparative assessments are no longer the most appropriate nor informative frame of reference for teachers to use in systems that have established a focus on learning outcomes and standards of expected achievements.

Central frameworks for summative reporting

As shown in Tables 4.A1.3, 4.A2.3a and 4.A2.3b (Annexes 4.A1 and 4.A2), almost all education systems for which information is available have policy frameworks that regulate the reporting of summative results. Such reporting frameworks tend to be developed at the central/state level and in most education systems they are compulsory for all schools[10]. The frameworks typically determine the frequency of summative reporting and the type of information to be provided to students and their parents. In Australia, for example, summative assessment practices are set at the state/territory level but there is a national requirement to report these in a nationally consistent way. Education systems vary regarding the frequency of formal reporting. In most systems, student results are reported in writing on average twice a year, but this may vary depending on the level of education.

The type of information provided in regular summative reports also varies across countries and levels of education. While the use of numerical, alphabetical or descriptive (e.g. excellent, good, satisfactory, unsatisfactory) marks is common at all levels of

education, it becomes more prevalent at the secondary level. In primary education, students are not awarded marks in Denmark, Iceland, New Zealand, Norway, Poland (Years 1-3 only), Slovenia (Years 1-3 only) and Sweden (Years 1-5 only). These countries rely instead on regular qualitative reporting, for example in the form of a summary of the student's strengths and weaknesses. Australia, Austria (Years 1-2 only), Finland, France, Ireland and Israel tend to use a mix of qualitative assessments and marks in primary education, whereas Hungary, Italy, Mexico, the Netherlands, Poland (Years 4-6), Slovenia (Years 4-6) and the Slovak Republic rely primarily on numerical marks for formal reporting. In Canada, requirements vary across provinces, and in Estonia, schools can decide which type of marking they use. At the lower and upper secondary level, all countries provide summary marks or ratings to students in regular reports. A number of countries, including Australia (general education only), France, Iceland (lower secondary only), Israel and Sweden (for students who fail in lower secondary only) complement these with qualitative written assessments.

In addition to requirements for regular written reporting, countries also frequently have central requirements for teachers to hold formal summative feedback meetings with students and their parents. Again, the most common periodicity for such meetings is twice a year, but there are variations across countries and levels of education. In France and Ireland, parents can request additional meetings beyond the mandatory meetings. In Spain, summative feedback meetings take place only at the request of parents. Denmark and Iceland have no particular requirements for such meetings in secondary education, and the Flemish and French Communities of Belgium, Finland, Hungary, Italy, Korea, New Zealand and Poland do not have such requirements at any level of education.

Finally, there are also variations across countries regarding the type of certification students receive at the end of key stages of education. At the end of primary education, just over half of the education systems for which information is available provide students with a formal certificate of learning (Tables 4.A1.3, 4.A2.3a and 4.A2.3b, Annexes 4.A1 and 4.A2). In most education systems where formal certificates are awarded, they contain a mark in each relevant subject at the end of primary education. In Finland, students usually also receive such report cards even though the national authorities do not require numerical marks before the end of Year 8. In France and Slovenia the subject marks are complemented by the overall average and a summary of the student's strengths and weaknesses (up to Year 3 in the case of Slovenia). In Poland, marks for behaviour, additional coursework and special achievements are also included. In Mexico, the formal certificate also includes Pass/Fail information and an overall mark. In Korea, the certificate only provides Pass/Fail information. In Northern Ireland in the United Kingdom, formal certificates of learning at ISCED levels 1 and 2 report the student's level of progression achieved in Communication and Using Maths (and in Using ICT from 2013/14).

At the lower secondary level, only the French Community of Belgium, Canada, the Czech Republic, the Slovak Republic and New Zealand do not award formal certificates of learning. In the majority of education systems, students receive a mark in each of the concerned subjects. In Finland, students receive an overall mark in addition to marks in each subject. Students receive Pass/Fail information in Korea, Mexico (in addition to an overall mark) and the Netherlands (in addition to a mark in each of the concerned subjects). In Denmark, in addition to subject marks, there is also a written assessment of a student project conducted in Year 9. Students also receive additional information in France (overall average and a summary of strengths and weaknesses), Iceland (qualitative assessments), Italy (overall mark based on an average of all tests including the national examination) and Poland (marks for behaviour, additional coursework and special

achievements). Some countries deliver a different type of certificate in pre-vocational and vocational programmes, such as the statement of attainment of competencies in Australia.

At the end of upper secondary education, all OECD education systems deliver a form of written certification, graduation report or diploma to students having completed the programme. In most systems, this takes the form of marks in each of the concerned subjects, overall marks or certain additional information.

Reported concerns about marking schemes

Marks awarded to students play an important role in all education systems. They may inform students and parents in a concise form about performance and return to efforts in learning processes. They can also potentially improve performance if they help to raise students' aspirations and convey that there are high returns to effort. Marks constitute an important source of information, as students rely on them to develop their expectations for further learning and enrolment in higher levels of education (OECD, 2012b). On the other hand, there are also risks that written marks may discourage effort and motivation if the information hurts self-confidence or convey to a student that returns to effort are low (Sjögren, 2009).

Also, across the countries participating in the OECD Review, there were a number of concerns regarding the ways in which marks were constructed and awarded to students. While most countries regulate the use of a particular marking scale (e.g. A-E or 1-10), especially in secondary education, this does not mean that the meaning of a particular mark is necessarily equivalent across schools. Even if schools use the same marking scale, they may have different marking criteria. It is difficult to ensure that the marks awarded in one school align with similar marks in another school. In addition, the same teacher will not necessarily be consistent in the application of criteria across students and over time. Such inequity in marking becomes problematic when a student moves to another school, or when marks are used for high-stakes decisions.

In many countries, teachers reported that overall marks were made up by assigning points to students across a range of elements including, for example, attendance, participation, homework completion, test performance, presentations, teamwork, neatness and discipline. However, the weight assigned to each of these elements was often not made explicit and varied across teachers, subjects and schools in the absence of rubrics specifying the meaning of points for each of these activities. Establishing marks by averaging points across a range of tasks and behaviours is likely to result in a kind of grand number with no clear significance for students (Ravela, 2009). Hence, it becomes difficult for students to understand what is expected of them, and obtaining a high mark may become the main objective for them with potentially distorting effects on learning efforts (Shepard, 2006). Parents may also become more concerned with the marks than about the actual learning progress of their children.

In several countries, there were concerns about a conflation of marks for performance and marks for behaviour. The practice of combining the assessment of behaviour with the assessment of actual achievement risks undermining the role of marks as indicators of performance. It may provide incentives for students to simulate effort as they learn how to behave with each teacher to make a good impression. Marks may also risk becoming a disciplinary instrument for teachers to control student behaviour rather than a tool to inform about learning and create a motivating learning environment (Shepard, 2006). While behaviour, effort and motivation are undeniably very important factors influencing student achievement, it would be preferable to report on these aspects separately. This

would allow communicating more accurate information about the student as a complex learner and provide indications about how particular learning needs can best be addressed. Education policy makers should consider providing guidance regarding appropriate reporting about student learning (see Box 4.11 for examples from Canada).

Box 4.11 Canada: Policies on reporting student performance

Similar to the policies of other jurisdictions in Canada, **Halifax** teachers are asked to develop assessment plans that are aligned with the Halifax Regional School Board of Nova Scotia *School Plan for Communicating Student Learning*. These plans must outline the purposes of the assessment mechanisms in light of the intended audiences and must be founded on individual learning practices accurately reflecting the learning outcomes as defined by the provincial curriculum and/or individual program plans. The policy indicates that comparisons of performance between students are to be avoided and that evaluation should not be based on measures such as students' social development and work habits, bonus points, student absence, missed/late assignments, group scores, neatness, etc. In addition, if report card marks are used, these should emphasise evidence collected from students' summative assessments. The policy also states that all actors (teachers and students) are called to interact continuously during the formative assessment process in order that information used to judge the progress leads to an understanding about the desired outcome and the criteria used to determine whether these outcomes have been successfully completed.

As for other Canadian jurisdictions, **Nova Central** School District's policy and procedures clearly indicates that a mark is not a simple reflection of a set of averages on a series of assignments but that it should be linked to a combination of evidences gathered from a variety of sources that students know and recognise. It is worth mentioning that teachers are directed not to look at the cumulative measures of achievement over time but that the final summative judgement be based on the student's most recent demonstration of his achievement of the desired outcome.

Source: Fournier and Mildon (forthcoming).

Reported concerns about communication with parents

Despite the existence of basic requirements regarding marking and certificates, in several countries there are no clear rules on how teachers should communicate assessment results to students and parents, and regular reporting practices are highly variable. In some settings, parents receive very limited feedback regarding their children's performance and progress. Throughout the OECD Country Review visits, parent representatives in different countries indicated that insufficient information was provided to them regarding the progress and identified needs for improvement of their children. In this context, several countries are working towards improving their reporting formats and harmonising reporting approaches across schools (for an example from Australia, see Box 4.12 below).

Absolum et al. (2009) criticise the very notion of "reporting" to parents as implying a power relationship that may inhibit meaningful partnership and dialogue. The term in fact suggests that parents are passive recipients of information rather than active partners of schools and supporters of their children's learning. The authors advocate for a mutual exchange of information between schools and homes where parents also share their understanding of their child's learning with the school. To establish such reciprocity, schools need to ensure that the information they provide is clear and easily understandable. Useful information, beyond simple marks, would include details about

students' progress, strengths, areas of need or concern, recommendations for further learning and illustrative examples.

Box 4.12 Innovations in the reporting system in Australia

In **Australia**, parental feedback at national, state and territory forums suggested that parents were confused by the different reporting scales and mechanisms used across schools. In response to these concerns, the Australian Government brought a degree of standardisation to teachers' judgements by requiring in 2005 that each state and territory adopt a common five-point scale as a condition for federal funding. At each year level from Year 1 to Year 10, teachers have to report students' achievements to parents using an A-E (or equivalent) framework. Defined specifically by each state and territory, generally the points on the A-E scale represent: A means well above standard; B means above standard; C means student at expected standard at time of report, on track to proficiency; D means below standard; E means well below standard.

A-E ratings are intended to assess the full range of learning expected of students. States and territories have developed guidelines and definitions for each of the A-E levels, variously labelled as letters (A-E) or descriptive categories (e.g. advanced – developing). States and territories vary in the specificity of the definitions and guidance they provide to support consistent judgements across teachers and schools. For example, Victoria provides teachers with detailed standards (*Victorian Essential Learning Standards*) co-ordinated with expected progression points, assessment maps and assessment modules to gauge student progress. Reporting software enables teachers to enter assessment scores and other ratings for components of each standard and the system automatically aggregates these scores into overall ratings for each student. To support consistency, Victoria also examines the relationship between the distribution of students' A-E ratings and NAPLAN results. There is a proposal to link A-E standards to the Australian Curriculum. The work has started in 2011 on a common approach to the achievement standards across states and territories including trialling and validation. In future, part of the work to align A-E with the Australian Curriculum will involve national agreement on definitions.

Source: Santiago et al. (2011).

Summative use of results

Cautions regarding potential misuse of summative assessment results

The results of summative assessment may be used by a range of stakeholders within and beyond schools, many of whom will have little or no training regarding measurement of student learning. Not all users of assessment results will be familiar with the intended uses of particular assessments, the content of the assessments, the evidence concerning the validity of inferences, and the characteristics of the test-takers. All of these information gaps may increase the likelihood of misuses of test results (Camara, n.d.). There is a large body of literature about standards for appropriate uses of assessment results. The *Standards for Educational and Psychological Testing* developed by AERA, APA and NCME (1999) are probably the most widely used reference in this domain.

Respecting a number of principles regarding the appropriate use of assessment results is essential to ensure responsible professional practice in educational assessment. Such principles include providing multiple opportunities for students to take alternate forms of an assessment, with opportunity for further learning between assessment administrations; drawing from multiple sources of information to enhance the validity of decisions being

made based on assessment results; considering the substitution of alternative measures for test scores, especially when tests are likely to give a deceptively low indication of a student's achievement because of test anxiety or disabilities that may reduce the validity of standardised test results; and ensuring that students have indeed had opportunities to learn the material that they are being assessed on (AERA, APA and NCME, 1999; Baker and Linn, 2004; Camara, n.d.).

Uses of summative assessment results across countries

Across OECD countries, the results of summative assessments are typically used for a wide range of decisions, including both within-school decisions and decisions beyond the school such as transfer to another school, transitions to higher levels of education and entry to the labour market. The remainder of this section provides an overview of country practices.

School-internal use of results

The primary function of summative assessment appears to be to keep records within schools and report to students, parents and other teachers. In most countries, summative assessment is conducted on a regular basis to collect information on student progress and check up on what students have retained and learned from a series of teaching units over a period of time (Harlen, 2007). This information is typically summarised in the form of marks and transcripts that students obtain at the end of course or year level. The aim most often is to provide a record of progress in learning to students, parents, other teachers and school leaders. To manage student assessment information on a longitudinal basis, some countries use computer applications to store student performance results, along with other information such as socio-demographic background and attendance.

Decisions on school admission, tracking and transfer

Schools may use summative assessments to decide on the admittance of students. This is more often the case in secondary than in primary schools. According to PISA 2009, on average across OECD countries, 36% of 15-year-old students are enrolled in schools whose principals reported that their schools are highly selective. In 10 of the 34 OECD countries, more than half of all 15-year-old students attend schools that always consider recommendations from feeder schools or academic transcripts when making admission decisions. Of these countries, in the Netherlands, Japan and Hungary, according to information provided by principals at the lower secondary level surveyed in PISA, more than 85% of students are selected for schools on the basis of academic records or recommendations. By contrast, less than 10% attend academically selective schools in Portugal, Spain, Iceland, Sweden, Finland, Denmark, Norway and Greece (OECD, 2010a).

Tracking into different school types

In a few OECD education systems, results are also used for tracking students into different school types at the end of primary education. In these cases, the assessment results of students in primary education have an effect on their transition from primary to secondary schools. In Europe, this is the case in Austria, Germany, Luxembourg and the Netherlands, and to some degree in the Flemish and French Communities of Belgium and Switzerland. In Austria, students need to have completed the fourth year of primary school with good marks in German and mathematics to be admitted to academic

secondary schools (*Gymnasium*). Students who are not admitted may take an entrance examination set by the *Gymnasium*. In Germany, after four years of primary education, primary schools give a recommendation on the type of lower secondary education the student should attend. Depending on the *Land*, the final decision is taken by the parents, the secondary school or the school authority. In Luxembourg, primary schools give a recommendation at the end of Year 6 in primary school, but students who are not admitted can take a national entrance examination. In the Netherlands, students' choice of secondary school is conditioned by their primary school's advice regarding the most appropriate programme of secondary education. This advice is based on the judgement of primary school teachers about the capacity of the student and on the student's results in a standardised but non-compulsory test in the final (sixth) year of primary education (Eurydice, 2009b).

Decisions on transferring students to another school

Assessment results may also be used to inform student transfers from one school to another. Transferring students to other schools because of low academic achievement, behavioural problems or special learning needs is a way for schools to reduce the heterogeneity of the learning environment. On average across OECD countries, at the lower secondary level, 18% of students attend a school in which school principals reported that the school would likely transfer students with low achievement, behavioural problems or special learning needs. Yet transfer policies vary across countries: in Iceland, Ireland, Portugal, Norway, Finland, the United Kingdom, New Zealand, Australia and Sweden less than 5% of students attend schools whose school principals reported that the school would likely transfer students for these reasons. By contrast, in Luxembourg, Austria, Belgium, Greece and Turkey, around one-third or more of students attend a school whose principal reported that students with low achievement, behavioural problems or special learning needs will "very likely" be transferred out of the school (OECD, 2010a).

Ability grouping within schools

Within schools, student assessment results may also be used to "stream" students into groups of similar ability levels. In PISA 2009, principals of lower secondary schools were asked to report whether students were grouped by ability into different classes or within a class, and whether these groupings were made in all or only selected subjects. Schools in most OECD countries use student assessment results in order to group students by ability within classrooms or between classrooms. On average across OECD countries, 13% of students are in schools whose principals reported that students are grouped by ability in all subjects, 55% are in schools whose principals reported that students are grouped by ability in some subjects and 32% are in schools whose principals reported that there was no ability grouping (OECD, 2010a).

Year repetition

In some countries, summative assessment results are used to inform decisions about year repetition (see also Chapter 3). This means that the results may be used to identify under-achieving pupils and decide on their progression from one year to the next. In many OECD countries, schools can decide to make students repeat a year if the assessment shows that they have not acquired adequate mastery of the curriculum at the end of a school year. However, there are large variations in country practices regarding the extent to which this possibility is used. In PISA 2009, on average 15% of 15-year-old students reported that they had repeated a year at least once: 7% had repeated a year in

primary school, 6% had repeated a year in lower secondary school and 2% had repeated an upper secondary Year. There are, however, marked differences across countries. While year repetition is non-existent in Korea, Japan and Norway, over 25% of 15-year-old students in France, Luxembourg, Spain, Portugal, Belgium and the Netherlands reported having repeated a year during their school trajectory in primary and secondary education (OECD, 2010a).

Decisions beyond school

In most OECD countries, a school leaving certificate is awarded to students after successful completion of compulsory education. In many OECD education systems, this certificate is at least partly based on results achieved in a final examination. But in some systems including Austria (up to 2013/14), the Flemish Community of Belgium, Finland, most German *Länder*, Hungary, Luxembourg and Turkey, the school leaving certificate is only based on teacher-assigned marks and work over the year. In several countries, the certificates are awarded on the basis of a combination between school-based marks and tests and external standardised exams (Eurydice, 2009b).

In all OECD countries, students receive a certificate after successful completion of upper secondary education. This certificate is generally a minimum requirement for admission to tertiary education. Only in Spain and Turkey the certificate is awarded only on the basis of continuous school-based assessment. In all other countries, certification is based at least partly on some kind of a final examination, often a combination of internal and external (or externally verified) exams. In France, Ireland and Slovenia, certification is based entirely on an external final exam (Eurydice, 2009b, updated with information from OECD Country Reviews).

In some education systems, the successful graduation from upper secondary school is a minimum requirement for entrance into higher education, for example in the Flemish Community of Belgium, France, Germany and the Netherlands. This does not mean that students are necessarily automatically admitted to higher education, but institutions of higher education and faculties may have their own supplementary entrance exams. Other countries such as Ireland, England, Korea and Australia directly use assessment for qualification and certification in upper secondary education for selection into higher education. In a further group of countries, upper secondary certification may serve a selection purpose only for subjects which are on high demand (Dufaux, 2012).

Even though the labour market does not have any minimum entrance requirements, students who do not choose to continue in higher education or another educational programme, but to apply for a job may undergo a similar process of selection. If the certificate of upper secondary education is perceived as a trustworthy institution for the communication of students' skills, employers may strongly base their selection of employees on the information provided through the certificate. Thus, certificates may have a strong screening or signalling function for employers. Reducing information asymmetry, employers may use certificates as a reliable tool to decide whether the applicant matches the requirements (Dufaux, 2012).

Formative use of results

As described above, the majority of countries participating in the OECD Review now have policy frameworks in place to support and promote formative assessment. At the same time, little comparable information is available about the way teachers actually document and use formative assessment results in practice.

Several countries have implemented measures for long- and medium-cycle formative assessment. This includes, for example, the use of Individual Development Plans (IDPs) in compulsory schools in Sweden. School leaders are required to set out the shape for the IDP. The Plans are to include an assessment of the student's current performance levels in relation to learning goals set in the curriculum and syllabi, and the focus is on steps that the student should take to reach those goals. Whether to include additional information, such as the student's more general development (e.g. the student's ability to take on responsibility, their social skills, and so on) is up to the school leader. For students who are experiencing difficulty, schools are required to document plans as to how they will help students achieve goals. The goals determined in IDPs are also used for student self-assessment in which students are asked to rate their own progress and performance. The IDP ensures that both teachers and students are focused on identifying individual learning goals, and developing strategies to address any shortcomings (Nusche et al., 2011b).

In Denmark, mandatory Individual Student Plans (ISPs) were introduced in 2006 to document student learning progress. According to Shewbridge et al. (2011), the ISPs contribute to formalising Danish assessment practice by documenting students' learning progress for dialogue with key stakeholders. They emphasise the student's future learning rather than summative learning outcomes. Official evaluations, strong support from national level parent organisations and student associations (see Danish Ministry of Education and Rambøll, 2011) and stakeholder feedback during the OECD Country Review confirm that the ISPs are well received by parents and teachers. In short, parents appreciate a written summary of their child's progress because they feel that they are better prepared for their meeting with teachers. Teachers perceive benefit in transferring documented information on student achievement to subsequent teachers and as such ISPs play a crucial role in tracking individual students' developmental growth over time. Teachers recognise the role of ISPs in easing communication with parents. The added workload ISPs entail for teachers is a bone of contention, but there is a current pilot to allow educators more flexibility in determining and prioritising the content of ISPs. Depending on the evaluation of this pilot, this may lead to a modified approach to drawing up ISPs (Shewbridge et al., 2011).

In the French Community of Belgium, the monitoring of students' progress towards core competencies is organised through individual learning plans (*Plans Individuels d'Apprentissage, PIA*). These plans are obligatory for students in special needs education, students with special learning needs who are integrated in mainstream education, as well as certain students facing important challenges in secondary education. The plan lists both subject-specific and transversal competencies to be acquired by students during a specified period. It is personalised in line with the potential and needs of each individual students and is regularly adjusted by the class council, based on observations provided by teachers and the school's guidance centre. The plans are used for formative purposes only, and are designed as much as possible in collaboration with the concerned students and their parents.

In Ireland, the National Council for Curriculum and Assessment (NCCA)'s assessment guidelines suggest that three records be maintained by primary schools: the teacher's day-to-day records, the pupil file and the report card. The pupil file is formative in purpose; it is used by teachers to record information on all aspects of the child's learning and development and provides a concise educational history of the child's ongoing progress. In Luxembourg, teachers are required to prepare formative reports (*bilans intermédiaries*) at the end of each trimester. These reports are descriptive in nature, do not contain test scores and are designed to maintain student motivation and

facilitate parents' understanding of student progress towards end-of-cycle objectives (Shewbridge et al., 2012).

Another long-cycle type of formative feedback is the organisation of regular development talks between school leaders/teachers and students and their guardians. In Norway, for example, teachers are expected to maintain documentation of their formative assessment of students and to meet with each student and his/her parents for a discussion of the student's progress once each term. In Sweden also, the IDPs with individualised goals are developed collaboratively in regular "development talks" between the teacher, individual students and their parents.

Such medium- and long-term formative uses of results are important for identifying areas for further improvement, developing broad teaching strategies to address needs identified within the student cohort, planning, allocation of resources, and so on. It can also feed into the school-wide coordination of pedagogical support and remediation for students facing learning difficulties. For example, Blondin and Giot (2011) describe a range of remediation strategies that were put in place in the French Community of Belgium to respond to student learning difficulties identified through formative assessment (Box 4.13).

Box 4.13 Organising remedial education and student orientation based on formative assessment results

In the **French Community of Belgium**, the organisation of formative assessment in all schools is mandated by decree. The importance of training teachers to be able to guarantee early identification of learning difficulties and immediate remediation is emphasised in the Community's Policy Declaration 2009-2014. While remediation strategies vary across schools, there are indications that schools are bringing together individual formative assessment results for the school-wide organisation of remediation activities. Several schools have established "needs-based groups" working on the basis of formative assessment results. Other schools have designated a teacher or member of school leadership to review requests for support, group them according to identified needs and ensure communication between classroom teachers and remedial teachers to provide coherent support to individual students. The decree also mentions the obligation for schools to organise student orientation by bringing together teacher teams, the psychological and social support centres (centres *psycho-médico-sociaux, CPMS*), parents and students. To this end, each school is serviced by a CPMS, whose multi-disciplinary teams offer free and confidential counselling. While the CPMS do not directly participate in the formative assessment of students, they do have the role of providing specific insights regarding adequate support and orientation strategies for students facing difficulties. They ensure coordination between pedagogical support in the classroom and broader psychological, medical, social and professional support for individual students.

Source: Blondin and Giot (2011).

While medium- and long-term strategies are important to ensure consistency of support throughout a student's learning trajectory, research indicates that short-cycle formative assessment – the daily interactions between and among students and teachers – has the most direct and measurable impact on student achievement (Looney, 2011a). In short-cycle interactions, formative assessment is part of the classroom culture, and is seen as an integrated part of the teaching and learning process. Teachers systematically incorporate such formative assessment methods in their course planning – for example, in how they intend to develop classroom discussions and design activities to reveal student

knowledge and understanding. These interactions encompass effective questioning to uncover student misconceptions and identify patterns in student responses, feedback on student performance and guidance on how to close learning gaps, and student engagement in self- and peer-assessment (OECD, 2005a; Wiliam, 2006; Looney, 2011a).

It is difficult to assess in how far teachers across countries are using this type of feedback in their regular classroom interactions with students. In several countries participating in the OECD Review, including those that already have firm foundations for formative assessments, there were concerns about the adequacy of regular teacher feedback. In several countries, despite some policy attention to formative assessment, feedback was conceived by teachers in a narrow way. It was frequently not immediate and tended to be quite limited in the form of marks or brief comments, focused on test results or performance rather than on learning. To be effective, formative assessment results should be used for feedback that is continuous and embodied in the process of teaching and learning itself (Wiggins, 1998).

Diagnostic use of results

Diagnostic assessment is a common form of formative assessment, which often happens as a kind of pre-assessment at the beginning of a study unit. Diagnostic assessment typically focuses on very specific areas of learning and produces fine-grained information about individual student strengths, weaknesses and learning needs. Many diagnostic tools are designed specifically to uncover the causes of students' learning difficulties. The results of diagnostic assessment are typically used to inform future programme planning, design differentiated instruction and deliver remedial programmes for at-risk students. The distinctive feature of diagnostic assessment, *vis-à-vis* formative assessment more generally, is its greater focus on the use of results for individualised intervention and/or remediation.

Pointers for future policy development

This chapter reviewed country approaches to the assessment of individual students in light of available research and evidence. Based on the analysis developed in this chapter, this section provides a range of policy options, which – collectively – have the potential to enhance student assessment frameworks across OECD countries. These pointers for policy development are drawn from the experiences reported in the Country Background Reports, the OECD Country Reviews and the available research literature.

It should be stressed that there is no single model or global best practice of student assessment. The development of practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing student assessment policies. In general, there is a need for further research into the impact of different policy approaches to student assessment. The existing evidence base is dominated by research in a few systems with long-established policies on student assessment. As more systems adopt and implement different student assessment policies, there will be a need to collect evidence on how these impact student learning and educational experiences.

Governance

The main conclusion in relation to governance is the need to develop a *coherent* framework for student assessment. Coherence implies that *(i)* the assessment framework is based on well-conceptualised reference documents (curricula, standards and learning progressions); *(ii)* the purposes of different assessment approaches are clearly set and complement each other; and *(iii)* the responsibilities for governing and implementing the assessment framework are well defined.

Establish a coherent framework for student assessment

Across countries there is increasing emphasis on designing and governing coherent assessment frameworks that integrate different types of assessments and use a range of information to make dependable judgements about student learning. Well-designed assessment frameworks can play a key role in building consensus about education goals, standards and criteria to judge proficiency. They can also be a lever to drive innovation in education by signalling the types of learning that are valued. Establishing clarity about the purposes and appropriate uses of different assessments is important to ensure that assessment frameworks optimally contribute to improvements at the classroom, school and system level. Building the assessment competencies of students, teachers and other stakeholders in the education system is crucial to ensuring the effective implementation of such frameworks.

A key governance challenge for countries is to develop a clear vision and strategy for assessment where different approaches developed nationally and locally each serve a clearly defined purpose and the format of the assessment is aligned to that particular purpose. For assessment to be meaningful, it must be well-aligned to the type of learning that is valued. For example, while simple knowledge tests are well-suited to assess the outcomes of traditional teaching approaches based on rote learning and knowledge transfer, such tests are less adequate when it comes to assessing complex competencies. Coherent assessment frameworks should aim to align curriculum, teaching and assessment around key learning goals and include a range of different assessment approaches and formats, along with opportunities for capacity building at all levels.

Develop clear goals and learning progressions to guide student assessment

In all student assessment systems, there is a need for clear external reference points in terms of expected levels of student performance at different levels of education. While it is important to leave sufficient room for teachers' professional judgements in the classroom, it is necessary to provide clear and visible guidance concerning valued learning outcomes in the central curriculum and standards. This is especially important as many curricula now highlight the need for students to acquire complex competencies. The challenge is that such competency goals are often stated in a general way with little guidance for teachers on how they can adapt their teaching and assessment to reach such goals. Hence, clear and concrete illustrations of the type of learning that should be achieved can provide important support.

Teachers can also benefit from specific learning progressions describing the way in which students typically move through learning in each subject area. Learning progressions can provide a picture from beginning learning to expertise and help provide teachers, parents and other stakeholders with concrete images of what to expect in student learning with direct links to the curriculum. Such learning progressions can provide a clear conceptual basis for a consistent assessment framework, along with assessment

tools that are aligned to different stages in the progressions. Clear descriptions and exemplars of expected learning, along with criteria to assess performance can provide further support.

Ensure a good balance between formative and summative assessment

A large body of research highlights the important role of formative assessment in improving learning outcomes. While most OECD countries have now developed policy frameworks to support and promote formative assessment, little information is available regarding the effective and systematic implementation across schools. There is a risk that pressures for summative scores may undermine effective formative assessment practices in the classroom. In fact, assessment systems that are useful for formative purposes are at risk of losing their credibility if high stakes are attached to them. Such tensions between formative and summative assessment need to be recognised and addressed.

Both formative and summative assessments should be well embedded within broader assessment frameworks. While summative assessment and reporting are important at key stages of the education process, it is the daily engagements of teachers' and students' with assessment information that will lead to sustainable improvements in learning outcomes. To support such practice, a national commitment to formative assessment on paper needs to be matched with a strategic approach to professional learning in assessment and concrete support for teachers and schools.

Establish safeguards against an overreliance on standardised assessments

A clear priority in assessment frameworks is the development of reliable measures of student learning outcomes. This effort has now started in most OECD countries with the development of standardised assessments in the main subjects at key stages of education. Standardised central assessments can help clarify learning expectations for all schools and motivate teachers and students to work towards high standards.

At the same time, there is a risk that the high visibility of standardised assessment, especially where high stakes are attached to them, might lead to distortions in the education process, such as excessive focus on teaching students the specific skills that are assessed, narrowing the curriculum, distributing repeated practice tests, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and potentially even manipulation of results. Because of these potential negative effects, it is important to establish safeguards against excessive emphasis on a particular standardised test and to draw on a range of assessment information to make judgements about learning progress.

Because standardised central assessment is a relatively new phenomenon in many OECD countries, it is important to be clear about its purposes, to develop large-scale assessments over time to be able to accommodate the purposes that are reasonable, point out inappropriate uses and provide guidance for the way in which these assessments can be used as part of a broader assessment framework. Assessment systems require research evidence on the extent to which the interpretations of assessment results are appropriate, meaningful, and useful. The role of the standardised assessments should be clearly fixed and the assessments should be continually developed, reviewed and validated to ensure that they are fit for purpose. Validation is a long-term process of accumulating, interpreting, refining and communicating multiple sources of evidence about appropriate interpretation and use of assessment information. Where new standardised assessments

are introduced, it is important that they are first trialled to enable an evaluation of impacts before full-scale implementation. It is also important for national authorities to clarify the kinds of decisions the assessments can provide evidence for and what decisions require other kinds of information.

Share responsibilities for the governance and implementation of assessment frameworks

Several actors are involved in governing and designing assessment frameworks, including different levels of the educational administration, specialised central assessment agencies, the inspectorates, private assessment developers, educational research centres, school leaders, teachers and students. To ensure the coherence of various assessment approaches, it is important that these different actors engage with each other and investments are made in leadership and collaboration around a joint assessment strategy. Responsibilities for the development and administration of assessments need to be shared between schools and external partners.

A balanced combination of teacher-based and external assessments would be most suitable to leverage both central expertise and teacher ownership and to ensure maximum validity and reliability. Learning outcomes that can be readily assessed in external examinations should be covered this way, whereas more complex competencies should be assessed through continuous teacher-based assessment. There are several ways to centrally support the quality and reliability of teacher-based assessment, for example through the use of scoring guides, negotiated scoring criteria, external benchmarks, training for teachers, multiple judgements and external moderation. It is also important to provide a range of nationally validated assessment tools that teachers can use to assess their students reliably when they see fit.

Procedures

The main conclusion in relation to procedures is the importance of developing a *comprehensive* set of assessment approaches. Comprehensiveness implies that the assessment framework uses a range of assessment instruments, formats and methods so that it captures the key outcomes formulated in national learning goals. The framework should allow teachers to draw on multiple sources of evidence in order to form dependable judgements on student learning. Comprehensiveness also means that assessment approaches are inclusive and able to respond to the various needs of *all* learners.

Draw on a variety of assessment types to obtain a rounded picture of student learning

A comprehensive assessment system should include a range of internal and external approaches and make use of different assessment formats to capture a broad range of learning outcomes for different purposes. It is not appropriate to try and serve multiple purposes with a single assessment. It is important, instead, to develop a comprehensive assessment system that is clear about what the various formats and approaches can achieve and ensures that they are used appropriately for their intended purpose. Providing multiple opportunities and formats for student assessment can increase both the validity and reliability of student assessment.

To ensure a broad approach to student assessment, education systems can provide a range of nationally validated assessment tools for different summative and formative purposes. In addition, teachers also need to build their competencies to develop valid and reliable assessment tools corresponding to specific local needs. Particular attention should be paid to ensuring that the breadth of curriculum and learning goals is maintained in student assessment by ensuring that all subject areas and objectives are given certain forms of attention.

Support effective formative assessment processes

While the importance of formative assessment is widely recognised across OECD countries, in many settings there is room for improving teachers' approaches to formative assessment. For example, formative assessment is sometimes understood by teachers as having many small practice tests in view of preparing a final summative assessment, or as providing feedback in the form of praise or encouragement to make more effort. However, for formative assessment to be effective it needs to be independent of the requirement to rate performance, and for feedback to be helpful for student learning, it needs to provide timely, specific and detailed suggestions on the next steps to enhance further learning.

Education authorities can support formative assessment procedures with a range of tools that may help schools in developing systematic approaches. The use of individual development plans for each student can support medium- and long-cycle formative assessment processes. In addition, specific guidelines, workshops, online tools and professional learning opportunities can support effective formative assessment on a daily basis (more on this under "Capacity").

Clarify and illustrate criteria to judge performance in relation to national goals

To assist teachers in their practical assessment work against learning goals, there is also a need to develop support materials, such as scoring rubrics listing criteria for rating different aspects of performance and exemplars illustrating student performance at different levels of achievement. Clear scoring rubrics can make teachers' assessment transparent and fair and encourage students' metacognitive reflection on their own learning. They can be used to define what constitutes excellent work and enable teachers to clarify assessment criteria and quality definitions.

Such guidance can help teachers make accurate judgements about student performance and progress, which is essential to make decisions about how to adapt teaching to students' needs. Teachers also need to acquire skills to develop their own specific objectives and criteria aligned with national learning goals, and should be encouraged to share and co-construct such assessment criteria with students, so that they understand different levels of quality work.

Ensure the consistency of assessment and marking across schools

While most countries set basic requirements regarding the use of particular marking scales and reporting formats, there tend to be large inconsistencies in marking practices across teachers and schools. Such inconsistency in approaches to marking reduces the value of marks as a tool to summarise and inform learning. It is also unfair to students, especially when marks are used to make decisions about their future educational trajectory. Central specifications regarding summative assessment and marking are important to help a consistent application of assessment criteria across schools.

In addition, moderation processes are key to increase the reliability of teacher-based assessment. Moderation involves strategies for quality assurance of assessment judgements, such as teachers cross-marking each other's assessments within a school or across schools, teachers discussing samples of student work in groups or in collaboration with experienced moderators, or a competent authority or examination board externally checking school-based assessments. The objective is to reduce variations in the ways teachers assess students and set marks in order to achieve fairness in student assessment and reporting.

Moderation practices should be encouraged for different types of assessments at all levels of education. It would be beneficial to develop guidelines and support for appropriate approaches to moderation, both within and across schools. Such guidelines should emphasise the importance of moderation as a process for developing assessment confidence and common understandings of assessment standards among teachers, but also as a mechanism to increase the dependability (validity and reliability) of teacher assessments of student performance.

Promote assessment formats that capture valued key competencies

Across OECD countries, there is increasing emphasis in curricula on the importance for students to acquire key 21st century competencies and education systems need to adapt their assessment approaches so that they promote and capture this broader type of learning. To this end, teachers need to be supported in translating competency goals into concrete lesson plans, teaching units and assessment approaches. Professional learning opportunities where teachers can discuss and collaborate in assessing actual student products can contribute to their understanding of broader assessment practices.

In addition, innovative assessment formats should also be developed centrally to complement teacher-made assessments. Due to concerns about reliability and resources, "performance-based" or "authentic" assessments are often challenging to implement on a large scale and in a standardised way. Alternatively, education systems can opt for developing test banks for teachers, which can provide a range of innovative assessment tools for teachers to draw from when their students are ready. Such test banks provide an excellent opportunity to promote innovative assessment tools that have proven successful elsewhere. They can offer a map of assessment items suitable to assess the key areas and competencies outlined in the curriculum.

Another option is to implement innovative assessments that cover larger parts of the curriculum on a sample basis. Sample-based assessments that are applied to a representative proportion of the student cohort allow the assessment of a broader range of curriculum content at relatively low cost while at the same avoiding distortions deriving from potential "teaching to the test". Such assessments may be organised in cycles, assessing a different curriculum area each year and not assessing all students on the same tasks, thus allowing the assessment of a wider range of content without overburdening individual students. While the purpose of such sample-based assessment typically is to monitor the education system (see Chapter 8), they can still be beneficial for individual teachers and students when they receive their results. The tasks of previous years may also be made available for teachers to use in their formative classroom assessment. Where teachers are centrally employed and trained to correct such sample-based assessments, this can constitute a valuable professional learning experience that will also help them in their classroom teaching and assessment practice.

Build on innovative approaches developed in particular education sectors

In many countries, there are some education sectors which have a longer tradition than others in using innovative assessment approaches. Often, there is a stronger tradition in the vocational education and training (VET) sector than in general education programmes to include innovative assessment approaches that may take place in practical and authentic work situations and are connected to real-life challenges that graduates may encounter in the workplace.

Sometimes there is also greater attention to assessing authentic performances in special needs education, second chance education programmes or special programmes for migrant students. In designing assessment approaches for general education programmes, it would be important to pay close attention to innovative assessments developed in other programmes and learn from approaches that have been shown successful and could be integrated and/or adapted. Policy makers should promote communication and collaboration regarding the assessment of competencies across education sectors and programmes, so that mutual learning can be facilitated.

Tap into the potential of ICT to develop sophisticated assessment instruments

Increasingly sophisticated ICT programmes that score open-ended performances, measure students' reasoning processes, examine how students go about thinking through problems and even provide feedback to students have been developed in some settings, predominantly in the United States. While it has always been possible for teachers or external assessors to perform these functions, ICT offers the possibility for large-scale and more cost-effective assessment of complex skills (Mislevy et al., 2001, in Looney, 2009).

While introducing constructed-response items and open-ended performance tasks in large-scale assessments is quite demanding, technology today makes this possible and more affordable. Increased efficiency would allow systems to administer tests, at different points in the school year, with results to be used formatively (as with curriculum-embedded or on-demand assessments). In addition, computer-based assessments can help increase the timeliness of feedback to teachers and students. While in many countries, central assessment systems provide teachers with results several months after the tests were administrated, the use of ICT-based assessment allows providing feedback to teachers very rapidly. With computer-based tests, it is possible to provide teachers and students with their test results the next day, which can foster the use of the test results for adapting teaching and learning for individual student progress.

In addition, computer-based assessments that adapt test items to the level of student performance on previous items can strengthen the diagnostic dimension of large-scale assessments. Only a few countries are using computer-based adaptive tests but these may provide powerful pedagogical tools for teachers. While a standardised test can only provide a snapshot of student achievement in selected targets and subjects, within a discrete area adaptive tests are able to provide a very accurate diagnosis of student performance. As each student sits a different test including questions that are adapted to his/her ability level, this can allow a more thorough diagnostic feedback.

These kinds of assessments are relatively new, and as of yet, relatively limited in number across OECD countries. However, as systems invest more in research to develop appropriate measurement technologies that are able to score complex performances and that reflect models of learning in different domains, development is likely to accelerate.

Ensure that student assessment is inclusive and responsive to different learner needs

Assessment systems should underline the importance of responding to individual learner needs and school community contexts, and design assessment strategies that suit the needs of different learner groups. The objective is to develop an inclusive student assessment system based on the principle that all students have the opportunity to participate in educational activities, including assessment activities, and to demonstrate their knowledge, skills and competencies in a fair way. Hence, teacher assessment practices as well as the content and format of standardised assessments should be sensitive to particular groups of students and avoid biases by socio-economic background, immigrant or minority status, and gender.

While innovative and motivating assessment strategies are important for all students, this is particularly the case for vulnerable students or students at risk of dropping out. Several studies indicate that certain formats of assessment may advantage or disadvantage certain student groups. Hence, to ensure fairness in assessment, it is important to offer a range of different assessment formats and tasks (e.g. test-based, performance-tasks, oral, written).

Dimensions of inclusive assessment, such as the sensitivity to cultural and linguistic aspects of assessment, should also be further included and developed in both initial education and professional development for teachers. The accessibility and lack of bias of standardised assessments for certain groups at risk of underachievement should receive due attention. This requires studies on differential test functioning for particular groups and the provision of specific test accommodations where necessary. It is suggested that quality assurance guidelines are prepared and practices adopted to ensure that assessments are reviewed for their potential bias in these respects.

Capacity

The main conclusion in relation to capacity relates to the need for assessment frameworks to be *participatory*. Student assessment involves a broad range of actors including students, teachers, school leaders, parents, education authorities and assessment agencies. All of these actors need to develop their competencies to ensure that stated objectives are reached, starting with the students themselves.

Put the learner at the centre and build students' capacity to engage in their own assessment

For assessment systems to enhance learning – and not just measure it – students need to be at the centre of the assessment framework. To become lifelong learners, they need to be able to assess their own progress, make adjustments to their understandings and take control of their own learning. Assessment can only lead to improvement in learning outcomes if students themselves take action and use assessment information to close gaps in their own learning. Recent educational research emphasises the importance of assessment as a process of metacognition, where learners become aware of their own thought processes, personally monitor what they are learning and make adaptations in their learning to achieve deeper understanding.

Self-and peer-assessment are powerful processes where students identify standards and criteria to make judgements about their own and their peers' work, which can promote a greater sense of agency and responsibility for their (life-long) learning. But

developing skills for self-assessment and self-regulation takes time and requires structured support by teachers in the classroom. Teachers can use classroom assessment to provide opportunities for students to engage in reflection and critical analysis of their own learning, for example by guiding students in setting learning goals and monitoring their progress towards them; working with them to develop criteria to judge progress; using exemplars and models of good practice and questioning of their own thinking and learning processes. Policy makers can support such practices by developing requirements, guidelines and support regarding learner-centred teaching and assessment.

Maintain the centrality of teacher-based assessment and promote teacher professionalism

Across many countries, there is recognition that teacher professionalism needs to be at the heart of effective assessment for learning. Students will develop their own assessment capacity only if teachers themselves have such capacity and are adequately resourced (Absolum et al., 2009). Placing a strong emphasis on teacher-based assessment has many advantages: it allows for competencies to be measured that are difficult to capture in a standardised assessment, it is embedded in regular coursework and more authentic than a test-based examination and it has greater potential to be used for subsequent improvements in teaching and learning.

However, in order to reach the full potential of teacher-based assessment, it is important for policy makers and stakeholders to adopt a strategic approach to teacher learning in assessment and invest in professional learning opportunities. Assessment capacity should be reflected in teacher standards and be addressed in a coherent way across teacher preparation programmes and publicly funded professional development programmes.

Teachers' assessment capacity can further be built and strengthened through systematic arrangements for moderation of assessments. There is considerable evidence that involving teachers in moderation is a powerful process not only for enhancing consistency but also for enabling teachers to deeply understand student learning objectives and to develop stronger curriculum and instruction. Moderated assessment and scoring processes are strong professional learning experiences that can drive improvements in teaching, as teachers become more skilled at various assessment practices and the use of assessment information to make adjustments to teaching and learning approaches.

Identify assessment priority areas for teacher initial education and professional development

There are variations across countries regarding the areas where teachers need most support to develop effective assessment practice. It is important for policy makers, together with teacher education institutions and stakeholders, to identify the topics related to assessment that are most in need of development within teacher education. Experience from the OECD Review indicates that the following areas require particular attention in many countries.

First, to be able to assess students' progress in developing complex competencies, it is important that teachers learn to develop a variety of assessment approaches and understand different aspects of validity, including what different assessments can and cannot reveal about student learning. Second, for summative teacher-based assessment to be reliable, it is important to provide focussed training on how to make summative

judgements on student performance in relation to national curriculum goals or standards and how to apply marking criteria to very different types of evidence of student learning. Third, for formative assessment to be effective, it is essential that teachers are offered in-depth professional learning opportunities in particular on embedding formative assessment in regular daily teaching practice, co-developing clear criteria for assessment with learners, giving specific, timely and detailed feedback, and creating conditions for students to develop self-monitoring skills. Fourth, to increase equity and fairness in assessment, training should also focus on ensuring that teachers are sensitive to cultural and linguistic aspects of learning and assessment.

Use teacher appraisal and school evaluation processes to help teachers develop their assessment capacity

Teacher appraisal and school evaluation processes can also contribute to identifying those areas of student assessment where teachers most need to develop their skills. Effective models of teacher appraisal and school evaluation specifically focused on teachers' and schools' assessment approaches have been piloted in some settings but are only in the early stages of development. Inspection visits, for example, may contribute to fostering innovation in assessment, by focussing on a wide range of quality indicators (beyond test results) including the capacity of teachers and schools to promote and assess key competencies.

Reporting and use of results

The main conclusion regarding the reporting and use of assessment results is the need for the assessment framework to be *informative*. It needs to produce high-quality information that can be shared with students, parents, school leaders and others with an interest in student learning outcomes. Reporting of assessment information needs to be clear, contextualised, and useful to foster learning and feed into decision making at different levels of the education system.

Develop clear reporting guidelines

Effective reporting is essential to communicate summary statements of achievement to students and their parents, as well as to other teachers within the school. Such records can support co-operation with parents, ensure consistency of support after student transitions to higher levels of education and provide a basis to make decisions about a student's further educational career. However, where there is a lack of transparency and consistency in the ways in which marks and report cards are constructed, the effect of such reporting will be counterproductive for student motivation and future learning.

Clear central reporting guidelines can help build a common understanding around the meaning of marks and the criteria used to establish student performance. They can also help to clarify that information about student behaviour, effort and motivation should not be mixed into performance marks. Keeping these elements separate in reporting allows communicating more accurate information about the student as a complex learner and can provide better indications about how particular learning needs can best be addressed.

Engage parents in education through adequate reporting and communication

Good reporting and communication strategies are important for involving parents in supporting their children's learning and in focussing resources, both at school and at home, on essential learning targets. Hence, reporting needs to be clear and easy to understand,

especially in primary education when parents and teachers can have the greatest impact on a child's learning. While some countries have standardised reporting approaches, others leave it to the local and school level to determine the format of reporting.

To ensure minimum quality requirements, countries could consider providing a template for reporting student achievement and provide guidance materials that teachers can use to report student performance in relation to student learning objectives. Useful information, beyond simple marks, would include details about student progress, strengths, areas of concern, identified needs, recommendations for further learning and illustrative examples.

Ensure transparency and fairness when using assessment results for high-stakes decisions

The results of summative assessment may be used by a range of stakeholders for different purposes and some of these users may have little knowledge of the intended uses of the assessments, the content of the assessments and the evidence concerning the validity of inferences from the assessments. Hence, there is a risk of misuse of assessment results. Also, several reviews of research have found the high stakes use of a single assessment to be strongly related with teachers focusing on the content of the assessments rather than underlying learning goals, administering repeated practice tests, training students for answering specific types of questions and students adopting surface learning techniques. To avoid such negative impacts on teaching and learning and reduce the risk of misuse of results, a number of principles on appropriate use of test results should be respected. For example, students should have multiple opportunities to show their learning, results from a single assessment alone should not be used to make high-stakes decisions about their future learning pathways, and alternative assessments should be considered to replace high-stakes testing, especially for students where there is a high probability that tests give a misleadingly low indication of their competencies.

Promote the regular use of assessment results for improvement

Assessment is closely intertwined with teaching and learning. Whether internal or external, assessment cannot be separated from a vision about the kind of learning that is valued and the teaching and learning strategies that can help students to get there. In turn, there is strong research evidence on the power of assessment to feed forward into new teaching and learning strategies and the strong relationship between assessment for learning and improvement of learning outcomes. To optimise the potential of assessment to improve what is at the heart of education – student learning – policy makers should promote the regular use of assessment results for improvements in the classroom. All types of assessment should have educational value, and be meaningful to those who participate in the assessment. To this end, it is important that all those involved in assessment at the central, local and school level have a broad vision of assessment and of the need to bring together results from different types of assessment activities to form rounded judgements about student learning and the use of assessment information for further improvement.

Notes

1. The use of assessment information to evaluate teachers, schools, school leaders and education systems will be discussed in Chapters 5 (teacher appraisal), 6 (school evaluation), 7 (school leader appraisal) and 8 (education system evaluation) of this report. Hence, in this chapter, standardised assessment is only included to the extent that it is used to make judgements about the learning of individual students – its school evaluation function is discussed in Chapter 6 and its national monitoring function is covered in Chapter 8. International student surveys will be discussed in Chapter 8.

2. See Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b in Annexes 4.A1 and 4.A2.

3. This refers to the state level in Australia, the Community level in the Flemish and French Communities of Belgium, the province/territory level in Canada and the regional government level in Northern Ireland.

4. In Italy, Luxembourg and Portugal, they apply to public schools only, in Austria they apply to public and government-dependent private schools only and in Denmark exception is granted to a few private schools.

5. Except in Italy, Ireland, Luxembourg and New Zealand, where they are not compulsory for all private schools.

6. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

7. The SAM-scale is available at the website *www.competento.be*. The website provides a non-exhaustive overview and links to screening tools for both education and non-education objectives (e.g. self-tests for students to evaluate their learning style). The website is supported by the Flemish Agency for Entrepreneurial Training (*Vlaams Agentschap voor Ondernemersvorming*).

8. The Educational Quality and Accountability Office of Ontario (EQAO), for example, has chosen option 2 and reports assessment results separately for French and English students. As results are not on the same scale, standard setting and reporting of results must be done separately. More information on EQAO testing in both English and French can be found on the EQAO website: *www.eqao.com/Parents/FAQ.aspx?Lang=E&gr=036*.

9. For instance, there could be a core of items or activities that would be the same, except for translation and adaptation. These items or activities would serve to equate results on other parts of the instrument which have not been translated and which are unique to each language group.

10. In Austria, Canada, Denmark (ISCED 1), Hungary (ISCED 2 and 3), Ireland and Mexico (ISCED 3), they are not compulsory for private schools.

References

Abedi, J. (2004), "The No Child Left Behind Act and English language learners: Assessment and accountability issues", *Educational Researcher*, Vol. 33, No. 1, pp. 4-14.

Abedi, J. and P. Gandara, (2006), "Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy", *Educational Measurement: Issues and Practices*, Vol. 25, No. 4, pp. 36-46.

Abedi, J., C. Hofstetter and C. Lord (2004), "Assessment accommodations for English language learners: Implications for policy-based empirical research", *Review of Educational Research*, Vol. 74, No. 1, pp. 1-28.

Absolum, M., L. Flockton, J. Hattie, R. Hipkins and I. Reid (2009), *Directions for Assessment in New Zealand (DANZ) Report: Developing Students' Assessment Capabilities*, Ministry of Education, Wellington.

ACACA (Australasian Curriculum, Assessment and Certification Authorities) (1995), *Guidelines for Assessment Quality and Equity*, ACACA, *http://acaca.bos.nsw.edu.au/files/pdf/guidelines.pdf*.

AERA, APA, NCME (American Educational Research Association, American Psychological Association, National Council for Measurement in Education) (1999), *Standards for Educational and Psychological Testing*, AERA, Washington D.C.

Alderson, C. and D. Wall (1993), "Does washback exist?", *Applied Linguistics*, Vol. 14, pp. 115-129.

Anderson, B., D.S. MacDonald and C. Sinnemann (2004), "Can measurement of results help improve the performance of schools?", *The Phi Delta Kappan*, Vol. 85, pp. 735-739.

Baartman, L.K.J., T.J. Bastiaens, P.A. Kirschner, C.P.M. Van der Vleuten (2006), "The wheel of competency assessment: Presenting quality criteria for competency assessment programs", *Studies in Educational Evaluation*, Vol. 32, pp. 153-170.

Baker, E.L. and R.L. Linn (2004), "Validity issues for accountability systems", in S.H. Fuhrman and R.F. Elmore (eds.), *Redesigning Accountability Systems for Education*, Teachers College Press, New York, pp. 47–72.

Barneveld, C. van (2008), "Using data to improve student achievement", *What Works? Research into Practice*, Research Monograph No. 15, The Literacy and Numeracy Secretariat, Ontario, *www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/Using_Data.pdf*.

Becker, W. and S. Rosen (1992), "The learning effect of assessment and evaluation in high school", *Economics of Education Review*, Vol. 11, No. 2, pp.107-118.

Bell, B. and B. Cowie (2001), *Formative Assessment and Science Education*, Kluwer Academic Press, Dordrecht, The Netherlands.

Bennett, R.E. (2001), "How the Internet will help large-scale assessment reinvent itself", *Education Policy Analysis*, Vol. 9, No. 5.

Biggs, J. (1999), *Teaching for Quality Learning at University*, SRHE and Open University Press, Buckingham, UK.

Biggs, J. (1998), "Assessment and classroom learning: A role for summative assessment?", *Assessment in Education: Principles, Policy & Practice*, Vol. 5, No. 1, pp. 103-110.

Biggs, J. (1996), "Enhancing teaching through constructive alignment", *Higher Education*, Vol. 32, pp. 347-364.

Binkley, M., O. Erstad, J. Herman, S. Raizen, M. Ripley with M. Rumble (2010), *Draft White Paper 1: Defining 21st Century Skills*, *http://atc21s.org/wp-content/uploads/2011/11/1-Defining-21st-Century-Skills.pdf*.

Bishop, J.H. (2006), "Drinking from the fountain of knowledge: Student incentive to study and learn – externalities, information problems and peer pressure", *Handbook of the Economics of Education*, Vol. 2, pp. 910-944.

Bishop, J.H. (1999), "Are national exit examinations important for educational efficiency?", S*wedish Economic Policy Review,* Vol. 6, No. 2, pp. 349-398.

Bishop, J.H. (1997), "The effect of national standards and curriculum-based examinations on achievement, *American Economic Review,* Vol. 87, No. 2, pp. 260-264.

Black, P. (1993), "Formative and summative assessments by teachers", *Studies in Science Education*, Vol. 21, pp. 49-97.

Black, P. and D. Wiliam (1998), "Assessment and classroom learning", *Assessment in Education: Principles, Policy and Practice,* Vol. 5, pp. 7-74.

Blondin, C. and B. Giot (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the French Community of Belgium,* Université de Liège, *www.oecd.org/edu/evaluationpolicy*.

Bloom, B.S. (1968), "Learning for mastery", *Evaluation Comment*, Vol. 1, pp. 1-12.

Bloom, B. (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*, by a committee of college and university examiners.

Boud, D. (1986), *Implementing Student Self Assessment*, Higher Education Research and Development Society of Australia (HERDSA), Sydney.

Boulet, M., G. Simard and D. Demelo (1990), "Formative evaluation effects on learning music", *Journal of Educational Research*, Vol. 84, pp. 119-125.

Brejc, M., M. Sardoč and D. Zupanc (2011), *Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Slovenia*, National School for Leadership in Education in cooperation with the National Examinations Centre, *www.oecd.org/edu/evaluationpolicy*.

Butler, R. (1988), "Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance", *British Journal of Educational Psychology*, Vol. 58, pp. 1-14.

Butler, D. and P.H. Winne (1995*),* "Feedback as self-regulated learning: A theoretical synthesis"*, Review of Educational Research*, Vol. 65, pp. 245-281.

Camara, W. (n.d), "Standards for Educational and Psychological Testing", The Gale Group, *www.education.com/partner/articles/galegrp/*

Chung, G.K.W.K. and E.L. Baker (2003), "Issues in the reliability and validity of automated scoring of constructed responses" in M.D. Shermis and J. Burstein (eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Lawrence Erlbaum, Mahwah, New Jersey, pp. 23-40.

Chung, G.K.W.K. et al. (2001), *Knowledge Mapper Authoring System Prototype*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Cizek, G. J. (1997), "Learning, achievement, and assessment: Constructs at a crossroads", in G. D. Phye (ed.), *Handbook of Classroom Assessment: Learning, Achievement, and Adjustment* (p 1-32), Academic Press, San Diego.

Clarke, M. (2012), "What matters most for student assessment systems: A framework paper", *SABER – Systems Approach for Better Education Results, Working Paper No. 1*, The World Bank, Washington, DC.

Clarke, M., W. Haney and G. Madaus (2000), "High stakes testing and high school completion", *National Board on Educational Testing and Public Policy Statement*, Vol. 1, No. 3, School of Education, Boston College, *www.bc.edu/research/nbetpp/publications/v1n3.html*.

Clarke, M., A. Shore, K. Rhoades, L. Abrams, J. Miao and J. Li (2003), *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States,* Lynch School of Education, Boston College, Massachusetts, National Board on Educational Testing and Public Policy.

Crooks, T.J. (2004), "Tensions between assessment for learning and assessment for qualifications", paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies (ACEAB), Fiji, March 2004.

Crooks, T.J. (1988), "The impact of classroom evaluation practices on students", *Review of Educational Research*, Vol. 58, pp. 438-481.

Danish Ministry of Education and Rambøll (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Denmark*, Agency for the Evaluation and Quality of Primary and Lower Secondary Education, *www.oecd.org/edu/evaluationpolicy.*

Darling-Hammond, L. (2010), *Draft White Paper 5: Policy Frameworks for New Assessments,* Assessment and Teaching of 21st Century Skills (ATCS), *http://atc21s.org/wp-content/uploads/2011/11/5-Policy-Framework.pdf.*

Darling-Hammond, L. (1998), "Policy and change: Getting beyond bureaucracy" in A. Hargreaves, A. Lieberman, M. Fullan and D. Hopkins (eds.), *International Handbook of Educational Change*, Kluwer Academic Publishers, Dordrecht.

Dee, T.S. and B.A. Jacob (2006), "Do high school exit exams influence educational attainment or labor market performance?", *NBER Working Paper 12199*, National Bureau of Economic Research, Cambridge, Massachusetts, *www.nber.org/papers/w12199*.

Dos Santos, S. and R. Rakocevic (2012), *Les cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude de l'OCDE : Rapport national de base pour la France*, Direction de l'Évaluation, de la Prospective et de la Performance (DEPP), Ministère de l'Éducation nationale, *www.oecd.org/edu/evaluationpolicy.*

Dufaux, S. (2012), "Assessment for qualification and certification in upper secondary education: A review of country practices and research evidence", *OECD Education Working Papers*, No. 83, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Earl, L. (2003), "Assessment as learning: Using classroom assessment to maximize student learning", in T.R. Guskey and R.J. Marzano (eds.), *Experts in Assessment Series*, Corwin Press, Thousand Oaks, California.

Earl, L. and M. Fullan (2003), "Using data in leadership for learning", *Cambridge Journal of Education*, Vol. 33, pp. 383-394.

Earl, L. and S. Katz (2008), "Getting to the core of learning: Using assessment for self-monitoring and self-regulation", in S. Swaffield (2008), *Unlocking Assessment: Understanding for Reflection and Application*, Routledge, Taylor and Francis, London.

Elmore, R. (1996), "Getting to scale with good educational practice", *Harvard Educational Review*, Vol. 66, No. 1.

Elmore, R. and G. Sykes (1992), "Curriculum policy", in P.W. Jackson (ed.), *Handbook of Research on Curriculum*, Macmillan, New York.

Elshout-Mohr, M., R. Oostdam and M. Overmaat (2002), "Student assessment within the context of constructivist educational settings", *Studies in Educational Evaluation*, Vol. 28, No. 4, pp. 369-390.

EPPI (2005), *Research Evidence on the Impact on Students of Self- and Peer-assessment*, Review conducted by the Assessment Review Group, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI), London.

EPPI (2004), *A Systematic Review of the Evidence of the Reliability and Validity of Assessment by Teachers for Summative Purposes*, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.

EPPI (2002), *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.

Eraut, M. (1990), "Identifying the knowledge which underpins performance" in H. Black and A. Wolf, *Knowledge and Competence: Current Issues in Education and Training*, Careers and Occupational Information Centre, Employment Department, London, UK.

European Commission (2012), *Assessment of Key Competences in Initial Education and Training: Policy Guidance*, Commission Staff Working Document, *http://ec.europa.eu/education/news/rethinking/sw371_en.pdf*.

European Commission (2011a), *Evidence on the Use of ICT for the Assessment of Key Competences*, European Commission, Brussels.

European Commission (2011b), *Assessment of Key Competences: Policy Handbook*, European Commission, Brussels.

Eurydice (2009a), *Key Data on Education in Europe 2009*, European Commission, Brussels.

Eurydice (2009b), *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results,* European Commission, Brussels.

Ewell, P.T. (2005), *Applying Learning Outcomes Concepts to Higher Education: An Overview*, prepared for the University Grants Committee.

Figlio, D.N. and M.E. Lucas (2004), "Do high grading standards affect student performance?", *Journal of Public Economics*, Vol. 88, pp. 1815-1834.

Finnish Ministry of Education and Culture (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Finland*, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, *www.oecd.org/edu/evaluationpolicy*.

Fournier, G. and D. Mildon (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Canada*, prepared for The Council of Ministers of Education, Canada (CMEC).

Frederiksen, J. and B. White (2004), "Designing assessment for instruction and accountability: An application of validity theory to assessing scientific inquiry", in M. Wilson (ed.), *Towards Coherence between Classroom Assessment and Accountability*, 103rd Yearbook of the National Society for the Study of Education, Part II, Chicago, Illinois.

Gagné, R.M. (1977), *The Conditions of Learning*, Holt, Rinehart & Winston, Inc., New York.

Gilmore, A. (2008), *Assessment Review Paper 8: Professional Learning in Assessment*, Ministry of Education, Wellington.

Gipps, C. and G. Stobart (2009), "Fairness in assessment", in C. Wyatt-Smith and J. Cumming (eds.), *Educational Assessment in the 21st Century*, Springer.

Gipps, C. and G. Stobart (2004), "Fairness in assessment", in *Perspectives on Pupil Assessment*, General Teaching Council for England (GTC), London.

Graham, A.E. and T.A. Husted (1993), "Understanding state variations in SAT scores", *Economics of Education Review,* Vol. 12, No. 3, pp. 197-202.

Greaney, V. and T. Kellaghan (1995), *Equity Issues in Public Examinations in Developing Countries,* World Bank, Washington DC.

Hajdúková, V., M. Šnídlová, A. Vykydalová, T. Lukács, A. Mentel, M. Pokorný, Z. Juščáková and R. Kanovská (forthcoming), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Slovak Republic,* National Institute for Certified Educational Measurements (NÚCEM), Bratislava, *www.oecd.org/edu/evaluationpolicy*.

Hamilton, L.S., B.M. Stecher, J. Marsh, J.S. McCombs, A. Robyn, J. Russell, S. Naftel and H. Barney (2007), *Implementing Standards-based Accountability Under No Child Left Behind: Responses of Superintendents, Principals, and Teachers in Three States,* RAND, Santa Monica, California.

Hargreaves, D.J., M.J. Galton and S. Robinson (1996), "Teachers' assessments of primary children's classroom work in the creative arts", *Educational Research*, Vol. 38, pp. 199-211.

Harlen, W. (2007), "Criteria for evaluating systems of student assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 15-28.

Harlen, W. (2005), "Teachers' summative practices and assessment for learning – tensions and synergies", *The Curriculum Journal*, Vol. 16, No. 2, pp. 207-223.

Harlen, W. (2004), "Can assessment by teachers be a dependable option for summative purposes?", *Perspectives on Pupil Assessment*, General Teaching Council for England (GTC), London.

Harlen, W. and M. James (1997), "Assessment and learning: Differences and relationships between summative and formative assessment", *Assessment in Education: Principles, Policy and Practice*, Vol. 4, No. 3, pp. 365-379.

Hattie, J. and H. Timperley (2007), "The power of feedback", *Review of Educational Research*, Vol. 77, No. 1, pp. 81-112.

Heritage, M. (2010), *Formative Assessment: Making It Happen in the Classroom*, Corwin Press, Thousand Oaks, California.

Herl, H.E. et al. (1999), "Reliability and validity of a computer-based knowledge mapping system to measure content understanding", *Computers in Human Behavior*, Vol. 15, pp. 315-334.

Herman, J.L., E. Osmundson and D. Silver (2010), *Capturing Quality in Formative Assessment Practice: Measurement Challenges*, CRESST Report 770, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Hidi, S. and J.M. Harackiewicz (2000), "Motivating the academically unmotivated: A critical issue for the 21st century", *Review of Educational Research*, Vol. 70, No. 2, pp. 151-179.

Hopper, M. (2001), *The Implications of Accommodations in Testing Students with Disabilities*, Educational Resources Information Center, *www.eric.ed.gov/ERICWebPortal/detail?accno=ED463627*.

Irish Department of Education and Skills (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Ireland, www.oecd.org/edu/evaluationpolicy*.

Irish Department of Education and Skills (2011*), Literacy and Numeracy for Learning and Life: The National Strategy to Improve Literacy and Numeracy among Children and Young People 2011-2020*, Dublin, *www.education.ie.*

Istance, D. and H. Dumont (2010), "Future directions for learning environments in the 21st century" in Dumont, H., D. Istance and F. Benavides (eds.), *The Nature of Learning*, OECD Publishing, Paris.

Jacob, B.A. (2005). "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools", *Journal of Public Economics*, Vol. 89, No. 5, pp. 761-796.

Kärkkäinen, K. (2012), "Bringing about curriculum innovations: Implicit approaches in the OECD area", *OECD Education Working Papers*, No. 82, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Kellaghan T., G. Madaus and A. Raczek (1996), *The Use of External Examinations to Improve Student Motivation,* American Educational Research Association (AERA), Washington DC.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and Jungha Park (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

Klein, S., L. Hamilton, D. McCaffrey and B. Stencher (2000), *What Do Test Scores in Texas Tell Us?,* Issue paper, Rand Education.

Klemp, G.O. (1977), "Three factors of success" in D.W. Vermilye (ed.), *Relating Work and Education*, Jossey-Bass, San Francisco, California.

Klenowski, V. (2009), "Australian Indigenous students: Addressing equity issues in assessment", *Teaching Education*, Vol. 20, No. 1, March 2009, pp. 77-93.

Koh, K. H., C. Tan, P. T. Ng (2012), "Creating thinking schools through authentic assessment: The case of Singapore", *Educational Assessment, Evaluation and Accountability*, Vol. 24, No. 2, pp 135-149.

Kohn, A. (1994), *The Risks of Rewards,* ERIC Clearinghouse of Elementary and Early Childhood Education, Illinois, US.

Kolb, D.A. (1981), "Learning styles and disciplinary differences", in A.W. Chickering (ed.), *The Modern American College*, Jossey-Bass, San Francisco, California.

Köller, O. (2001), "Mathematical world views and achievement in advanced mathematics: Findings from the TIMSS population III", *Studies in Educational Evaluation*, Vol. 27, pp. 65-78.

Koretz, D., R.L. Linn, S.B. Dunbar and L.A. Shepard (1991), "The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests", paper presented at the annual meeting of the American Educational Research Association (AERA), Chicago, 11 April 1991.

Lewis, A. (1998), *Assessing Student Achievement: Search for Validity and Balance*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Linn, R.L. (2000), "Assessments and accountability", *Educational Researcher*, Vol. 29, No. 2, pp. 4–16.

Linn, R.L., E.L. Baker and S.B. Dunbar (1991), *Complex, Performance-Based Assessment: Expectations and Validation Criteria*, CSE Technical Report 331.

Looney, J. (2011a), "Integrating formative and summative assessment: Progress toward a seamless system?", *OECD Education Working Papers*, No. 58, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Looney, J. (2011b), "Alignment in complex systems: Achieving balance and coherence", *OECD Education Working Papers*, No. 64, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Looney, J. (2009), "Assessment and innovation in education", *OECD Education Working Papers*, No. 24 , OECD Publishing, Paris, *www.oecd.org/edu/workingpapers.*

Lucas, B. and G. Claxton (2009), *Wider Skills for Learning: What are they, How can they be Cultivated, How could they be Measured and Why are they Important for Innovation?*, Centre for Real-World Learning.

Luedemann, E. (2011), "Intended and unintended short-run effects of the introduction of central exit exams: Evidence from Germany", in E. Luedemann, *Schooling and the Formation of Cognitive and Non-cognitive Outcomes*, Chapter 4, ifo Beiträge zur Wirtschaftsforschung 39, ifo Institut, Munich.

Madaus, G.F. and L.M. O'Dwyer (1999), "A short history of performance assessment: Lessons learned", *The Phi Delta Kappan*, Vol. 80, No. 9, pp. 688-695.

Marsh, C.J. and G. Willis (2007), *Curriculum: Alternative Approaches, Ongoing Issues*, Pearson Education Inc., New Jersey.

Marzano, R.J. (2001), *Designing a New Taxonomy of Educational Objectives*, Corwin Press.

Masqud, M. and C. Pillai (1991), "Effect of self-scoring on subsequent performances in academic achievement tests", *Educational Research,* Vol. 33, pp. 151–154.

McCurdy, B. and E. Shapiro (1992), "A comparison of teacher monitoring, peer monitoring, and self monitoring with curriculum based measurement in reading among students with learning disabilities", *Journal of Special Education,* Vol. 22, pp. 162–180.

McDonald, B. and D. Boud (2003), "The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations", *Assessment in Education*, Vol. 10, No. 2, pp. 209–220.

MCEETYA (2008), *Melbourne Declaration on Educational Goals for Young Australians*, Ministerial Council on Education, Employment, Training and Youth Affairs, Canberra.

McFarlane, A. (2003), "Editorial. Assessment for the digital age", *Assessment in Education*, Vol. 10, pp. 261–266.

McGehee, J.J. and L.K. Griffith (2001), "Large-scale assessments combined with curriculum alignment: Agents of change", *Theory into Practice*, Vol. 40, pp. 137-144.

McMurrer (2007), *Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era*, Center on Education Policy, *www.cep-dc.org/displayDocument.cfm?DocumentID=312*.

Merrett, J. and F. Merrett (1992), Classroom management for project work: An application of correspondence training, *Educational Studies*, Vol. 18, No. 1, pp. 3-10.

Mischo, C. and F. Rheinberg (1995), "Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung", *Zeitschrift für Pädagogische Psychologie*, Vol. 9, pp. 139-151.

Mislevy, R.J., L.S. Steinberg, F.J. Breyer, R.G. Almond and L. Johnson (2001), *Making Sense of Data From Complex Assessments*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Morris, A. (2011), "Student standardised testing: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 65, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Natriello, G. (1987), "The impact of evaluation processes on students", *Educational Psychologist*, No. 22, pp. 155-175.

Nayral de Puybusque, A. (forthcoming), "Equity in student assessment: A literature review", *OECD Education Working Papers*, OECD Publishing, Paris.

Newton, P.E. (2007), "Clarifying the purposes of educational assessment", *Assessment in Education*, Vol. 14, No. 2, pp. 149-170.

New Zealand Ministry of Education New Zealand Ministry of Education (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for New Zealand, www.oecd.org/edu/evaluationpolicy.*

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, *www.oecd.org/edu/evaluationpolicy.*

Nusche, D. (2008), "Assessment of learning outcomes in higher education: A comparative review of selected practices", *OECD Education Working Papers*, No. 15, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Nusche, D. (forthcoming), "Summative assessment: What's in it for students?", *OECD Education Working Papers*, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

OECD (2012a), *Education at a Glance 2012: OECD Indicators*, OECD Publishing, Paris.

OECD (2012b), *Grade Expectations: How Marks and Education Policies Shape Students' Ambitions*, OECD, Paris.

OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, Paris.

OECD (2010a), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

OECD (2010b), *Learning for Jobs*, OECD Reviews of Vocational Education and Training, OECD Publishing, Paris.

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS,* OECD Publishing, Paris.

OECD (2007), *Qualifications Systems: Bridges to Lifelong Learning*, OECD Publishing, Paris.

OECD (2005a), *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD Publishing, Paris.

OECD (2005b), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OECD Publishing, Paris.

Ou, D. (2010), "To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam", *Economics of Education Review*, Vol. 29, pp. 171-186.

Papay, J.P., R.J. Murnane and J.B. Willett (2008), "The consequences of high school exit examinations for struggling low-income urban students: Evidence from Massachusetts", *Working Paper 14186*, National Bureau of Economic Research, Massachusetts, *www.nber.org/papers/w14186.*

Pascarella, E.T. and P.T. Terenzini (2005), *How College Affects Students. A Third Decade of Research*, Vol. 2, Jossey-Bass, San Francisco, California.

Pellegrino, J.W., N. Chudowsky and R. Glaser (2001), *Knowing What Students Know*, National Academy Press, Washington, DC.

Pepper, D. (2011), "Assessing key competences across the curriculum – and Europe", *European Journal of Education*, Vol. 46, No. 3.

Popham, J. (1991), "Why standardized tests don't measure educational quality", *Educational Leadership,* Vol. 56, No. 6, pp. 8-15.

Posner, G.J. (1992), *Analyzing the Curriculum*, McGraw-Hill, US.

Pryor, J. and H. Torrance (1998), "Formative assessment in the classroom: Where psychological theory meets social practice", *Social Psychology of Education*, Vol. 2, No. 2, pp. 151-176.

Ravela, P. (2009), "Consignas, devoluciones y calificaciones: los problemas de la evaluación en las aulas de educación primaria en América Latina", Páginas de Educación, Vol. 2, Universidad Católica del Uruguay, Montevideo, pp. 49-89, *www.ucu.edu.uy/Default.aspx?tabid=1267*.

Rawlins, P., J. Brandon, J. Chapman, L. Leach, G. Neutze, A. Scott and N. Zepke (2005), "Standards-based assessment in the senior secondary school: A review of the literature", *New Zealand Journal of Teachers' Work*, Vol. 2, Issue 2, pp. 107-115.

Reynolds, C.R. (1998), "Cultural bias in testing of intelligence and personality", in C. Belar (ed.), *Comprehensive Clinical Psychology*: *Sociocultural and Individual Differences*, Elsevier Science, New York.

Ripley, M. (2009), "Transformational computer-based testing", in F. Scheuermann and J. Björnsson (eds.), *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing,* Office for Official Publications of the European Communities, Luxembourg.

Rivera, C., E. Collum, L. Schafer Willner and J.K. Sia Jr. (2006), "Analysis of state assessment policies regarding the accommodation of English language learners", in C. Rivera and E. Collum (eds.), *State Assessment Policy and Practice for English Language Learners,* Lawrence Erlbaum Associates, Mahwah, New Jersey.

Rosenkvist, M.A. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Ross, J.A., C. Rolheiser and A. Hogaboam-Gray (1999), "Effects of self-evaluation training on narrative writing", *Assessing Writing*, Vol. 6, pp. 107-132.

Rothman, R. (2010), *Principles for a Comprehensive Assessment System*, Policy Brief, Alliance for Excellence in Education, Washington, DC.

Rupp, A.A. and N.K. Lesaux (2006), "Meeting expectations?: An empirical investigation of a standards-based assessment of reading comprehension", *Educational Evaluation and Policy Analysis*, Vol. 28, No. 4, pp. 315-333.

Rychen, D.S. (2004), *Key Competencies for All: An Overarching Conceptual Frame of Reference*, DeSeCo Strategy Paper on Key Competencies, *www.unal.edu.co/diracad/competencias/KeyCompetences.pdf*.

Sadler, R. (1989), "Formative assessment and the design of instructional systems", *Instructional Science*, Vol. 18, pp. 119-144.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012b), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012c), *OECD Reviews of Evaluation and Assessment in Education: Mexico*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Sattler, J.M. (1992), *Assessment of Children*, 3rd Edition, Jerome M. Sattler, San Diego.

Sawyer R., S. Graham and K. Harris (1992), "Direct teaching, strategy instruction and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities", *Journal of Educational Psychology,* Vol. 84, pp. 340–350.

Scheerens, J., M. Ehren, P. Sleegers and R. de Leeuw (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Netherlands,* University of Twente, *www.oecd.org/edu/evaluationpolicy*.

Scriven, M. (1967), "The methodology of evaluation", *AERA Monograph Series on Evaluation*, Vol. 1, pp. 39-83.

Sebba, J. and G. Maxwell (2005), "Queensland, Australia: An outcomes-based curriculum", in *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD Publishing, Paris, pp. 191-204.

Shavelson, R.J., G.P. Baxter and J. Pine (1990), "What alternative assessments look like in science", paper presented at the Office of Educational Research and Improvement Conference: The Promise and Peril of Alternative Assessment, Washington, October 1990.

Shepard, L. (2006), "Classroom assessment", in R. Brennan (ed.), *Educational Measurement*, ACE/Praeger, Westport, pp. 623-646.

Shepard, L.A. and K. Cutts-Dougherty (1991), "Effects of high-stakes testing on instruction", paper presented at the annual meeting of the Educational Research Association.

Shewbridge, C., M. Ehren, P. Santiago and C. Tamassia (2012), *OECD Reviews of Evaluation and Assessment in Education: Luxembourg,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shultz, K.S. and D.J. Whitney (2005), *Measurement Theory in Action*, Sage Publications, Thousand Oaks.

Sims, D. (2008), "Strategic responses to school accountability measures: It's all in the timing", *Economics of Education Review*, Vol. 27, No. 1, pp. 58-68.

Singh, M. (2003), "Understanding life skills", background paper prepared for the Education for All Global Monitoring Report 2003/4 "Gender and Education for All: The Leap to Equality", UNESCO Institute for Education, Hamburg, *http://unesdoc.unesco.org/images/0014/001469/146963e.pdf*.

Sjögren, A. (2009), "The long run consequences of being graded in elementary school", paper presented at the Final Conference of the RTN Network "Economic of Education and Education Policy in Europe" (EEEPE), hosted by the Centre for Economic Performance, LSE, London.

Slomp, D. (2008), "Harming not helping: The impact of a Canadian standardized writing assessment on curriculum and pedagogy", *Assessing Writing*, Vol. 13, No. 3, pp. 180-200.

Somerset, A. (1996), "Examinations and educational quality", in A. Little and A. Wolf, *Assessment in Transition, Learning Monitoring and Selection in International Perspective*, Pergamon, Elsevier Science Ltd., Oxford.

Specht, W. and F. Sobanski (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Austria*, Bundesminiterium für Unterricht, Kunst und Kultur (BMUKK), Vienna, *www.oecd.org/edu/evaluationpolicy*.

Stecher, B. and S. Barron (2001), "Unintended consequences of test-based accountability when testing in 'Milepost' grades", *Educational Assessment,* Vol. 7, No. 4, pp. 259–81.

Stiggins, R. (2005), "From formative assessment to assessment for learning: A path to success in standards-based schools", *Phi Delta Kappan*, Vol. 87 (4), pp. 324-328.

Stiggins, R.J. (1999), "Assessment, student confidence and school success", *Phi Delta Kappan*, Vol. 81, No. 3, pp. 191-198.

Stiggins, R.J. (1987), "Design and development of performance assessments", *Education Measurement: Issues and Practice*, Vol. 6 No. 3, pp. 33-42.

Stiggins, R. and J. Arter (2007), *Classroom Assessment for Student Learning: Doing it Right – Using It Well,* Pearson Education, Toronto.

Stiggins, R.J., M.M. Griswold and K.R. Wikelund (1989), "Measuring thinking skills through classroom assessment", *Journal of Educational Measurement*, Vol. 26, pp. 233-246.

Swaffield, S. (2008), "Feedback: The central process in assessment for learning", in S. Swaffield (ed.) *Unlocking Assessment: Understanding for Reflection and Application*, Routledge, Taylor and Francis Group, London and New York.

Thompson, S.J., A.B. Morse, M. Sharpe and S. Hall (2005), *Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of Students with Disabilities*, The Council of Chief State School Officers, *www.ccsso.org/Documents/2005/Accommodations_Manual_How_2005.pdf*.

Thurlow, M.L., L.L. Christensen and K.E. Lail (2008), "An analysis of accommodations issues from the standards and assessments peer review", *NCEO Technical Report 51*, National Center on Educational Outcomes, *www.cehd.umn.edu/nceo/OnlinePubs/Tech51/Technical51.pdf*.

Timperley, H. (2011), *Realising the Potential of Professional Learning*, University Press, Berkshire.

Timperley, H., A. Wilson, H. Barrar and I. Fung (2008), *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration (BES),* New Zealand Ministry of Education, Wellington.

van de Vijver, F. and N.K. Tanzer (2004), "Bias and equivalence in cross-cultural assessment: An overview", *Revue Européenne de Psychologie Appliquée*, No. 54, pp. 119-135.

Volante, L. (2008), "Equity in multicultural student assessment", *Journal of Educational Thought*, Vol. 42, No. 1, pp. 11-26.

Volkwein, J.F. (2003), "Implementing outcomes assessment on your campus", *Research and Planning EJournal*, Vol. 1, No.1.

Wandall, J. (2010), *National Tests in Denmark: CAT as a Pedagogical Tool*, Danish National School Agency, Copenhagen.

Wang, J. and M.S. Brown (2007), "Automated essay scoring versus human scoring: A comparative study", *The Journal of Technology, Learning, and Assessment"*, Vol. 6, No. 2, *www.jtla.org*.

Westbury, I. (2007), "Making curricula: Why do states make curricula, and how?" in M. Connelly, M.F. He and J.A. Phillion (eds.), *Handbook of Curriculum and Instruction*, Sage, Thousand Oaks, California.

Whiting, G. and D. Ford (n.d.), *Cultural Bias in Testing*, The Gale Group, *www.education.com/reference/article/cultural-bias-in-testing/*.

Wiggins, G. (1998), *Educative Assessment – Designing Assessments to Inform and Improve Student Performance*, Joey-Bass, San Francisco.

Wiliam, D. (2006), "Formative assessment: Getting the focus right", *Educational Assessment*, Vol. 11, pp. 283-289.

Wiliam, D., C. Lee, C. Harrison and P.J. Black (2004), "Teachers developing assessment for learning: Impact on student achievement", *Assessment in Education*, Vol. 11, pp. 49-65.

Wiliam, D. (2010), "The role of formative assessment in effective learning environments", in Dumont, H., D. Istance and F. Benavides (eds.), *The Nature of Learning*, OECD, Paris.

Woessmann, L., E. Luedemann, G. Schuetz and M.R. West (2009), *School Accountability, Autonomy and Choice around the World*, Edward Elgar, Cheltenham, UK.

Annex 4.A1 Features of student assessment frameworks in lower secondary education

The tables below provide information on features of student assessment frameworks in lower secondary education in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Part of the information was supplied by countries through a questionnaire specifically developed by the OECD Review. The rest of the information is based on data provided by OECD *Education at a Glance 2011*.

Additional material on features of student assessment frameworks in primary and upper secondary education is available on line at *http://dx.doi.org/10.1787/9789264190658-8-en*.

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

General notes

Australia: Australia is a federation of eight states and territories. There are differences in policy frameworks for evaluation and assessment across states and territories as well as between public (government) and private (non-government) schools.

Belgium (Fl., Fr.): In Belgium, education policy is the responsibility of each Community. The terms "national" and "central", therefore, refer to the highest educational authorities (Ministries of Education) of the Flemish and French Communities of Belgium.

Belgium (Fl.): For public schools, the school organising bodies are typically the central educational authority (Flemish Community) and provincial/regional and local educational authorities (provinces, cities, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Belgium (Fr.): For public schools, the school organising bodies (education networks) are typically the central educational authority (French Community) and provincial and local educational authorities (provinces, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Canada: Canada comprises ten provinces and three territories. Provincial/territorial education authorities refer to the highest level of educational authorities in Canada, as there is no federal/central department of education. There are differences in policy frameworks for evaluation and assessment across provinces and territories.

Chile: For public schools, the school organising bodies (sustainers) are typically local educational authorities (municipalities).

Czech Republic: For public schools, the school organising bodies are typically local education authorities at ISCED levels 1 and 2 and regional education authorities at ISCED level 3.

Hungary: For public schools, the school organising bodies (maintainers) are typically local and regional educational authorities. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

Ireland: School boards of management comprise members external to the school such as representatives of the patron and of the local community as well as members internal to the school such as the principal, teacher representatives and parent representatives.

Netherlands: In principle, all schools are government funded. Depending on their denomination, they can have a private (religious or pedagogy-based) or public character. For public schools, school organising bodies (competent authorities) can be local educational authorities (municipal authorities), a local governing committee with transferred powers of the municipality, or a public foundation or corporation. School organising bodies for private schools can be groups of parents, foundations or corporations.

New Zealand: School Boards of Trustees typically comprise elected members from the school community, the principal, a staff representative and a student representative (in secondary schools).

Spain: Responsibilities for education are shared between the central educational authority (Ministry of Education, Culture and Sport [*Ministerio de Educación, Cultura y Deporte*]) and state educational authorities (Regional Ministries or Departments of Education of the Autonomous Communities [*Comunidades Autónomas*]). The central educational authority executes the general guidelines of the government on education policy and regulates the basic elements or aspects of the system. The Autonomous Communities develop the central regulations and have executive and administrative competences for managing the education system in their own territory. State educational authorities refer to educational authorities at the highest level of the Autonomous Communities. Throughout the tables, the Autonomous Communities are referred to as "state educational authorities".

United Kingdom (Northern Ireland): Following political agreement to the devolution of certain policy and legislative powers from the United Kingdom government at Westminster to a local Assembly in 1999, legislative responsibility for education in Northern Ireland was one of the functions devolved to the Assembly and to a locally elected Minister for Education. The Department of Education, Northern Ireland, provides the central governance and management of education in Northern Ireland and is referred to as the "central level" throughout the tables.

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012)

This table describes central/state policy frameworks for internal summative assessment at ISCED level 2, i.e. forms of student assessment that are not standardised, but designed and marked by students' own teachers and implemented as part of regular classroom instruction in schools. This table focuses on summative forms of internal assessment, i.e. assessment designed to provide a summary statement of student achievement at a particular point in time in order to record, mark or certify achievements that may count towards a student's year-end or final marks. Summative assessment usually occurs at the end of a learning unit, term, school year, or educational level.

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Australia	General	Yes, at the state level for all schools	State agency	State curriculum goals; state standards[1]	Moderation of marking; availability of state guidance materials for marking student performance on the examination	100%
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)	State agency	National standards	Moderation of marking; availability of state guidance materials for marking student performance on the examination	100%
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	Central education authority or government	National curriculum goals	Competences for marking are included in initial teacher education	100%
Belgium (Fl.)	All programmes	No, but certain basic requirements are set at the central level for all schools[2]	Inspectorate	National curriculum goals (attainment targets and developmental objectives)	None	m
Belgium (Fr.)	All programmes	No, but certain basic requirements are set at the central level for all schools[3]	Inspectorate	National curriculum goals	Availability of guidance materials for marking student performance in the examination (provided either by education authorities or school organising bodies [education networks])	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Canada	All programmes	No, but certain basic requirements are set at the provincial/territorial level for public schools and government-dependent private schools only	Provincial/territorial education authorities or governments and a provincial agency in one province (Ontario)	Varies across provinces/territories	Varies across provinces/territories	0%
Chile	All programmes	Yes, at the central level for all schools[4]	School (school principal and teachers)	National curriculum goals; national standards	None	100%
Czech Republic	All programmes	No, but certain basic requirements are set at the central level for all schools[5]	School board or committee; Czech School Inspectorate (CSI)	National curriculum goals; national standards	None	100%
Denmark	All programmes	Yes, at the central level for all schools[6]	Central agency (National Agency for Quality and Supervision)	Binding national objectives (year-level objectives and end objectives for compulsory education)	Availability of national guidance materials for marking student performance on the examination (performance criteria)	50%
Estonia	General	Yes, at the central level for all schools	School (school principal)	National curriculum goal; national standards	None	100%
	Pre-voc and voc	No	a	a	a	a
Finland	All programmes	Yes, at the central level for all schools[7]	Teachers and school leaders	Final assessment criteria in each subject as part of the national core curriculum[8]	Availability of national performance criteria for the transition points in each subject	Not specified in the policy framework
France	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Hungary	All programmes	Yes, at the central level for all schools	School organising bodies (maintainers) and school leaders	National curriculum standards	None	100%
Iceland	All programmes	No, but certain basic requirements are set at the central level for all schools[9]	School board	National curriculum goals	None	m
Ireland	All programmes	Yes, at the central level for public and government-dependent private schools only	Teachers and school leaders	National curriculum goals	Availability of national guidance material for assessing student performance[10]	100%
Israel	All programmes	No	a	a	a	a
Italy	All programmes	Yes, at the central level for all schools[11]	School board or committee (includes the school principal and all teachers)	National curriculum goals[11]	None	100%
Korea	All programmes	Yes, at the central level for all schools	Respective subject committee within the school	National curriculum goals	Decision made by the School Mark Management Committee within the school	100%
Luxembourg	All programmes	Yes, at the central level for public schools	Central education authorities or government; school principal	National curriculum goals	None	100%
Mexico	All programmes	Yes, at the central level for all schools	State education authorities or governments	National curriculum goals	None	100%
Netherlands	All programmes	Yes, at the central level for all schools	School organising bodies (competent authorities)	National curriculum goals	Moderation of marking (within schools); Inspectorate monitors differences between school-based assessment and central examinations at the end of the cycle	End of year: 100%; End of cycle: 50%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
New Zealand	All programmes	Yes, at the central level for state schools and state-integrated schools only	School Board of Trustees	National curriculum goals	Availability of national guidance materials for marking student performance (performance criteria, exemplars); moderation of marking (within and between schools)	100%
Norway	All programmes	No, but certain basic requirements are set at the central level for all schools[12]	Local education authorities	National curriculum goals	None	100%
Poland	All programmes	Yes, at the central level for all schools[13]	School principal	National standards; curriculum decided at the school level	None	100%
Portugal	General	Yes, at the central level for public schools only	National education authority; pedagogical council	National curriculum goals	Moderation of marking within schools (criteria approved by the pedagogical council)	70%
	Pre-voc and voc	No	a	a	a	a
Slovak Republic	General	No, but certain basic requirements are set at the central level for all schools	School principal	National standards, national education programme	None	m
	Pre-voc and voc	No	a	a	a	a
Slovenia	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Spain	General	Yes, at the central level for all schools[14]	State inspection bodies	National curriculum goals	Availability of national guidance materials for marking student performance on the examination (performance criteria)	100%
	Pre-voc and voc	No	a	a	a	a
Sweden	All programmes	No	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	Yes, at the central level for all schools (Department of Education, Northern Ireland)	Council for Curriculum, Examinations and Assessment (CCEA)	Statutory rule on Levels of Progression which is to be laid before the Northern Ireland Assembly	There is both internal moderation within schools and moderation by CCEA	Not specified in the framework

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

1. Australia: A national curriculum (The Australian Curriculum) is currently being implemented and is expected to replace state-based curriculum frameworks.

2. Belgium (Fl.): At present, the Ministry of Education and Training does not write comprehensive curricula. However, the Ministry specifies attainment targets and developmental objectives (*eindtermen en ontwikkelingsdoelen*), which function as the basis for each curriculum. The attainment targets and developmental objectives are also called the "core curriculum". They are set by the Flemish central education authority. It is mandatory for schools to monitor their educational quality, which implies that each school must at least assess the extent to which students acquire the skills, knowledge and attitudes described in the attainment targets and developmental objectives. In this respect each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of its students. The Inspectorate may ask schools to present their policy. To some extent, internal summative assessment at schools is regulated by the attainment targets. Schools are also required to assess whether students have attained the objectives of the school curriculum to a satisfactory level. However, there are no regulations regarding the methods schools can use to do so.

3. Belgium (Fr.): The policy framework is provided through the work of the Inspectorate and the possibility for students to appeal summative assessment decisions made by the class council through an appeals council.

4. Chile: The policy framework sets very basic requirements, such as the number of assessments per year. It also states that every school has to define an internal protocol for summative assessment.

5. Czech Republic: The central education authority does not determine the framework for the internal summative evaluation. However, the Education Act requires every school to set its own assessment system that is approved by the school board. Internal school assessment regulations and their administration are controlled by the Czech School Inspectorate (Česká školní inspekce [CSI])

6. Denmark: A few private schools may be exempt from these requirements.

7. Finland: To some extent, the national core curriculum functions as a policy framework that regulates internal assessments.

8. Finland: In basic education, the marking scale ranges from 4 (weak) to 10 (excellent) and the curriculum contains descriptions of "good" performance (mark 8) at the so-called transition points for every subject. At the end of basic education there are final assessment criteria for the mark 8. Teachers hold responsibility for the design of assessments, but design assessments in relation to these descriptions and criteria. The tests are, therefore, based on the same criteria.

9. Iceland: At completion of compulsory school (Year 10), students undergo a summative assessment. It is, however, at the discretion of schools themselves how this is implemented and may take the form of a formative assessment. Requirements stipulated in legislation, regulations and curriculum guidelines are flexible.

10. Ireland: Guidance material is available from syllabus assessment guidelines, the State Examinations Commission and other sources. Syllabus and curricular guidelines, developed by the National Council for Curriculum and Assessment, provide assessment supports, as does material provided by the Professional Development Service for Teachers. The State Examinations Commission provides guidance for teachers in summative assessment through the publication of examination papers, marking schemes and chief examiners' reports in different subjects.

11. Italy: Presidential Decree n. 122/2009 provides norms on student assessment. The national curriculum goals are not explicitly mentioned in this regulation, but they are normally considered by school committees.

12. Norway: There are regulations in the Education Act that state that students at this level shall receive summative overall achievement marks in all subjects in addition to ongoing formative assessment and formal formative feedback minimum twice a year. The overall achievement marks based on teachers judgements are entered on the leaving certificate in Year 10.

13. Poland: While a policy framework exists, it is very general and leaves a great degree of autonomy to schools to define their own assessment rules. Schools are, however, required by law to define assessment regulations as part of the school statute.

14. Spain: Subject teachers are responsible for the assessment and marking of stud+A38ents at this level. There are, however, some constraints and guidelines. Minimum subject contents, learning objectives (both summative and formative) and assessment criteria are established at a central level. At the school level subject departments write a general programme for the year, revised by the inspection services to make sure it complies with the legal framework, establishing the assessment and marking methods and criteria. Subject departments are also responsible for making a written test (optional for those students who have previously attained the objectives of the school year), common to all school groups of the same level for year-end assessment and marking. Finally, the class teachers board has a say in the end of year assessment, especially at the end of each cycle and in relation to the attainment of the general objectives of the stage.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A1.2 Student formative assessment frameworks at ISCED levels 1, 2 and 3 (2012)

This table describes central/state policy frameworks for student formative assessment, i.e. assessment that aims to identify aspects of learning as it is developing in order to deepen and shape subsequent learning rather than make a judgement about the past performance of the student. It is essentially a pedagogical approach consisting of frequent, interactive checks of student understanding to identify learning needs and adapt teaching strategies.

Country	Is there a policy framework for promoting student formative assessment in the classroom?	What requirements are part of the policy framework for promoting student formative assessment in the classroom?
	1	2
Australia	ISCED levels 1, 2 and 3 (general): Yes, at the state level for all schools ISCED level 3 (pre-voc and voc): No	ISCED level 1: For schools to implement strategies for student formative assessment ISCED levels 2 and 3: For schools to implement strategies for student formative assessment; for student formative assessment to be part of initial teacher education programmes
Austria	No	a
Belgium (Fl.)	No[1]	a
Belgium (Fr.)	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment
Canada	Varies across provinces/territories	Varies across provinces/territories
Chile	No	a
Czech Republic	No	a
Denmark	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment[2]
Estonia	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment; for schools to report on their strategies to promote student formative assessment
Finland	No[3]	a
France	No	a
Hungary	No	a
Iceland	ISCED levels 1 and 2: yes, at the central level for all schools[4] ISCED level 3: no	For schools to implement strategies for student formative assessment
Ireland	ISCED level 1: yes, at the central level for public schools and government-dependent private schools only ISCED levels 2 and 3: no[5]	For schools to implement strategies for student formative assessment
Israel	ISCED levels 1 and 2: yes, at the central level for all schools ISCED level 3: no	For schools to implement strategies for student formative assessment
Italy	Yes, at the central level for all schools[6]	For schools to implement strategies and criteria for student formative assessment
Korea	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment; for student formative assessment to be part of initial teacher education programmes; for teachers to undertake professional development in student formative assessment

Table 4.A1.2 Student formative assessment frameworks at ISCED levels 1, 2 and 3 (2012) (*continued*)

Country	Is there a policy framework for promoting student formative assessment in the classroom?	What requirements are part of the policy framework for promoting student formative assessment in the classroom?
	1	2
Luxembourg	Yes, at the central level for public schools only	For schools to implement strategies for student formative assessment
Mexico	ISCED 1 and 2: yes, at the central level for all schools ISCED 3: yes, at the central level for centrally managed public schools, at the state level for locally managed public schools[7]	For schools to implement student formative assessment; for student formative assessment to be part of initial teacher education programmes[8]
Netherlands	No[9]	a
New Zealand	Yes, at the central level for state schools and state-integrated schools only	For schools to implement strategies for student formative assessment
Norway	Yes, at the central level for all schools[10]	For schools to implement student formative assessment
Poland	ISCED levels 1, 2 (general), ISCED level 3: yes, at the central level for all schools ISCED level 2 (pre-voc and voc): No	For schools to implement student formative assessment[11]
Portugal	ISCED 1, 2 (general) and 3 (general): yes, at the central level for all schools	For schools to implement strategies for student formative assessment (approved by the pedagogical council at each school)
Slovak Republic	No	a
Slovenia	Yes, at the central level for all schools	For schools to implement student formative assessment
Spain	ISCED levels 1, 2 (general) and 3: yes, at the central level for all schools ISCED level 2 (pre-voc and voc): no	For schools to implement student formative assessment; for schools to report on their strategies to promote student formative assessment; for student formative assessment to be part of initial teacher education programmes; for teachers to undertake professional development in student formative assessment
Sweden	No[12]	a
United Kingdom (Northern Ireland)	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment

Notes: a: information not applicable because the category does not apply; pre-voc and voc: pre-vocational and vocational programmes.

1. Belgium (Fl.): At ISCED level 1, schools are required to monitor the progress of every student and report their observations to parents. Schools are, however, not restricted by any regulations on how to implement progress monitoring or reporting. All relevant agreements on monitoring and assessment have to be included in the school development plan and in the school regulations. At the start of the school year, every parent signs the school regulations for agreement. At ISCED levels 2 and 3, the "counselling class committee" has a formative purpose. The counselling class committee is obliged to monitor the students' performance and assess their progress. The committee is staffed by teachers who teach the students involved. The school principal, deputy principal, technical advisor, members of the Centre for Pupil Guidance (*CLB*) and others may also be asked to participate in meetings of the guiding class committee, e.g. to give advice. According to central government regulations the guiding class committee is not qualified to decide whether a student passes or fails a school year or to take decisive disciplinary measures like refusing students to attend school during the next school year. This decision is made by the deliberative class committee. The counselling class committee must perform the following actions on a regular basis: 1) Analyse and discuss the students' school results as well as their attitudes; 2) Provide students with proper counselling if necessary, e.g. as soon as learning difficulties have been diagnosed; 3) Reduce failure rates by advising students to alter their course of study or – for future reference – to determine the course of study that fits them most.

2. Denmark: Individual Student Plans are compulsory in all subjects at least once a year for Years 1 to 7. For Years 8 and 9 individual student plans are combined with student learning plans.

3. Finland: The Finnish national core curricula for all ISCED levels mention that teachers should observe student's progress, but not how they should do so. It is at the discretion of local education authorities or individual teachers to decide measures for formative assessment.

4. Iceland: There has been increased emphasis on formative assessment *vis-à-vis* summative assessment. A new policy framework for student assessment is being developed with the publication of the revised curriculum guidelines in 2012.

5. Ireland: For ISCED levels 2 and 3 (general programme only), the National Council for Curriculum and Assessment (NCCA), a central agency, has issued guidelines to schools that emphasise the value and uses of formative assessment (Assessment for Learning). These guidelines are not in the form of regulations or statutes. The new Framework for ISCED 2 to be introduced in 2014 will provide for the promotion of student formative assessment in the classroom.

6. Italy: Central legislation requires schools to set strategies and criteria for formative assessment, but the procedures for doing so are within the scope of school autonomy (Law 122/2009, Artt. 1, 2, 3, 4).

7. Mexico: At schools managed by autonomous agencies (public and private), policies for student formative assessment in the classroom at ISCED level 3 are determined at the school level by autonomous and private institutions (e.g. universities).

8. Mexico: This information reflects the Comprehensive Reform of Basic Education (*Reforma Integral de la Educación Básica* [RIEB]) and the Comprehensive Reform of Upper Secondary Education (*Reforma Integral de la Educación Media Superior* [RIEMS]).

9. Netherlands: There are no formal frameworks for the formative assessment of students. However, there are formative assessment systems (e.g. by the Central Institute for Test Development (*Centraal Instituut voor Toetsontwikkeling* [CITO]) in use in primary and, increasingly, secondary education. Also, draft laws are being prepared that will require schools to use formative assessment systems for results-based work in schools.

10. Norway: Principles of formative assessment are part of the regulations in the Education Act.

11. Poland: The central regulations are very flexible and leave the choice of assessment systems to schools. They only prescribe the format of marks used at the end of the school year. They require the formative use of school assessment by defining the formative use of assessment as one of the main functions of student assessment.

12: Sweden: It is at the discretion of local education authorities or individual teachers to decide measures for formative assessment.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012)

This table describes central/state policy frameworks for the reporting of students' summative results to students and parents at ISCED level 2. Summative results in this table comprise both students' results in standardised national examinations and internal summative assessments.

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Australia	General	Yes, at the central level for all schools; Yes, at the state level for all schools[1]	Varies nationally (on average twice a year)	Alphabetical marks; descriptive ratings; qualitative assessment	Varies nationally (on average twice a year)	Mark in each of the concerned subjects
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)[1, 2]	Twice a year	Descriptive ratings; alphabetical mark	Twice a year	Statement of attainment of competencies
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	Twice a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Belgium (Fl.)	All programmes	No[3]	a	a	a	a
Belgium (Fr.)	All programmes	No	a	a	a	a
Canada	All programmes	Yes, at the provincial/territorial level for public schools and government-dependent private schools only	In most cases once at the end of the school year	Varies across provinces/territories	Varies across provinces/territories	No formal certificate of learning is awarded
Chile	All programmes	Yes, at the central level for all schools	Typically 2-3 times a year	No requirement exists	Typically 2-3 times a year	Primary certificate and marks in each of the concerned subjects
Czech Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Denmark	All programmes	Yes, at central level for all schools	At least twice a year	Numerical marks (-2 to 12)	Once a year	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in student project in Year 9

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Estonia	General	Yes, at central level for all schools	At least twice a year	Schools can decide whether to provide numerical marks or alphabetical marks	Once a year	Mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Finland	All programmes	Yes, at the central level for all schools[4]	Twice a year	Requirement to report numerical marks (4-10) from Year 8 onwards	No requirement exists	Overall mark; mark in each of the concerned subjects
France	All programmes	Yes, at the central level for all schools	At least 3 times a year	Numerical marks (0-20); descriptive ratings of competencies; qualitative assessment	At least three times a year; at the request of parents	Mark in each of the concerned subjects; summary of the student's overall strengths and weaknesses; progression to the next year level; grade point average
Hungary	All programmes	Yes, at the central level for public schools only	Twice a year	Numerical marks (1-5)	No requirement exists[5]	Mark in each of the concerned subjects
Iceland	All programmes	Yes, at the central level for all schools for Year 10	Twice a year	Qualitative assessment; numerical marks (0–10)	No requirements exists	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in each subject
Ireland	General	No, but advice and guidance are provided by a central agency (National Council for Curriculum and Assessment [NCCA])[6]	Typically twice a year	Typically numerical/ alphabetical marks and descriptive ratings	Typically once a year; Additional meetings facilitated at the request of parents/teachers	Typically alphabetical mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Israel	All programmes	Yes, at the central level for all schools[7]	Twice a year	Descriptive ratings; qualitative assessment; numerical marks	Twice a year	Mark in each of the concerned subjects

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Italy	All programmes	Yes, at the central level for all schools[8]	2-3 times a year	Numerical marks (1-10)	No requirement exists[9]	Mark in each of the concerned subjects; overall mark (provided in ISCED 2 certification at the end of Year 8 based on average of all tests including the national examination)
Korea	All programmes	Yes, at the central level for all schools	4 times a year	Numerical marks (0-100) and rankings	No requirement exists	Pass/fail information
Luxembourg	All programmes	Yes, at the central level for all schools	3 times a year	Numerical marks (0-60 points), descriptive ratings[10]	At the request of students/parents only[11]	Mark in each of the concerned subjects
Mexico[12]	All programmes	Yes, at the central level for all schools	More than 4 times a year	Numerical marks (5-10)	More than 4 times a year	Pass/fail information; overall mark (provided in the *Certificado de Terminación de Estudios*)
Netherlands	All programmes	Yes, at the central level for all schools	3 times a year	Numerical marks (0-10)	3 times a year	Pass/fail information; mark in each of the concerned subjects
New Zealand	All programmes	No	a	a	a	a
Norway	All programmes	Yes, at the central level for all schools	Once a year	Numerical mark (1-6)	Twice a year	Mark in each of the concerned subjects
Poland	General	Yes, at the central level for all schools	Twice a year	Descriptive ratings;[13] mark for behaviour/conduct	No requirement exists	Descriptive ratings;[13] mark for behaviour/conduct; comments on additional coursework and special achievements
	Pre-voc and voc	No	a	a	a	a
Portugal	General	Yes, at central level, for all schools	3 times a year	Numerical marks (1-5)	3 times a year and/or at request of parent	A formal certificate of learning is awarded when requested; mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Slovenia	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Slovak Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Spain	General only	Yes, at the central level for all schools[14]	3 times a year	Numerical marks (0-10)	At the request of parents	Mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Sweden	All programmes	Yes, at the central level for all schools	Twice a year	Descriptive ratings; qualitative assessments	Twice a year	Mark in each of the concerned subjects
United Kingdom (Northern Ireland)	All programmes	Yes, at the central level for all schools	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[15]	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[15]

Notes: a: information not applicable because the category does not apply; pre-voc and voc: pre-vocational and vocational programmes.

1. Australia is a federation of eight states/territories. Standardised examinations and internal summative assessment practices are set at the state/territory level and there is a national requirement to report student results in a nationally consistent way (A-E reporting).

2. Australia: Vocational education and training (VET) qualifications approved under the Australian Qualifications Framework (AQF) at ISCED levels 2 and 3 are typically offered by the VET sector. Students may take VET qualifications as part of a general programme (e.g. through the programme VET in Schools). VET (AQF) qualifications are delivered by a Registered Training Organisation (RTO) or a school in partnership with a RTO. All schools (public and private) are able to undertake partnerships with RTOs to provide students with an opportunity to undertake VET courses.

3. Belgium (Fl.): In secondary education, the "deliberative class committee" decides whether or not students pass or fail. The deliberative class committee's decisions rely on concrete data from the file of the student. The "counselling class committee" has a formative purpose. The counselling class committee is obliged to monitor the students' performance and assess their progress. The information in primary and secondary education can be provided as a qualitative assessment or by numerical marks (overall mark and marks/subject), depending on the school policy. At the end of Years 2 and 4 of lower secondary education, students receive a certificate.

4. Finland: The information provided refers to the reporting of summative results through report cards/certificates. Assessment is based on the national core curriculum. Standardised national examinations are only held at the end of ISCED level 3 (general programme).

5. Hungary: At the discretion of each school.

6. Ireland: Advice and guidance are provided by the National Council for Curriculum and Assessment (NCCA), but not in form of a statute.

7. Israel: The information provided refers mainly to a report card at the end of the school year.

8. Italy: From age 16 onwards, ISCED level 2 qualifications can also be obtained from provincial centres for adult education.

9. Italy: Decisions on formal summative feedback meetings with students/parents rest within schools, per law n. 122/2009. Typically meetings take place 2-3 times a year.

10. Luxembourg: For details, see the following website: *www.men.public.lu/priorites/ens_fondamental/090723_bibliotheque/111201_intermediate_reports_cycle2.pdf*.

11. Luxembourg: Each secondary school is required to organise an information meeting once a year (1st or 2nd trimester). However, there is no obligation to provide summative feedback as part of these meetings.

12. Mexico: From 2012-13 onwards, the new general norms on evaluation, accreditation, promotion and certification in basic education will apply. The Basic Education Certificate (*Certificado de Educación Básica*) will be issues on completing Year 12.

13. Poland: 6: Excellent, 5: Very good, 4: Good, 3: Satisfactory, 2: Acceptable, 1: Unsatisfactory.

14. Spain: More complete reports are given only to students with special educational needs or when a learning problem arises. This typically involves the identification of learning difficulties and related advice (e.g. year repetition, curricular adaptation, etc.).

15. United Kingdom (Northern Ireland): From 2013/14, the levels of progression achieved in Using ICT will also be provided.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central examinations

Austria: The introduction of national examinations at ISCED level 3 starts in school year 2014/15 for general programmes, and in school year 2015/16 for pre-vocational and vocational programmes. A central agency, the Federal Institute for Research, Innovation and Development of the Austrian School System (*Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens* [BIFIE]) is responsible for the development of the central examination. The following subjects will be examined in a standardised form at ISCED level 3: the language of instruction (German, Croatian, Hungarian, Slovenian), foreign languages (English, French, Spanish, Italian, Latin, Greek) and mathematics. Schools/teachers will be responsible for marking the new national examination from 2014/15 onwards following guidelines developed by BIFIE.

Belgium (Fr.): Standardised central examinations at ISCED levels 2 and 3 are not mandatory. Schools decide about their administration. At ISCED level 2, 55.6% of students participated in these examinations in 2011/12. At ISCED level 3, standardised national examinations in general programmes examine one competency area in history, and standardised examinations in pre-vocational and vocational programmes examine one competency area (reading) in the language of instruction (French). Schools decide about the weight given to standardised examination results in relation to other results of the student. In 2011/12, 42.9% of students participated in these examinations.

Iceland: With the introduction of the Compulsory School Act 2008, standardised national examinations were changed to standardised national assessments.

New Zealand: Students can participate in national qualifications examinations at any stage and it is not uncommon for some students to participate at ISCED level 2. However, the majority of the students undertake the national qualifications at ISCED level 3.

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012)

This table describes central/state policy frameworks for standardised central examinations, i.e. standardised student assessments that have a formal consequence for students (e.g. impact upon a student's eligibility to progress to a higher level of education, part of a process to certify learning) for ISCED level 2.

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Australia	All programmes	No	a[1]	a	a	a	a	a	a	a	a	a	a
Austria	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fr.)	General	Yes	*CE1D (Épreuve certificative externe commune au terme de la troisième étape du continuum pédagogique)*	Yes, for public schools only (from 2012/13)	All students: M, L	Year 8	Central education authority	School[2]	Availability of national guidance materials for marking student performance on the examination	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National standards
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Canada	All programmes	Yes, at the provincial level	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province
Chile	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Czech Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Denmark	All programmes	Yes	*Folkeskolens Afsluttende Prøver*	Yes, for public schools only[3]	All students: M, L, S, FL Students choose: A, V Sample: FL, S, SS, R	Year 9	Central education authority	Centrally appointed external examiners mark student performance in co-operation with internal examiners	Availability of national guidance materials for marking student performance on the examination (performance criteria, exemplars, rubrics, keys); moderation of marking (external examiners attend all oral examinations at school level)	Multiple choice; Open-ended writing tasks; oral presentation; oral questions and answers; project presentation; group discussion (optional)	Closed-format short-answer questions, open-ended writing tasks	No	National learning progressions

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Estonia	General	Yes	Basic school end exams	Yes, for public and government-dependent private schools only	All students: M; L Students choose: S; SS; FL	Year 9	State education authorities or governments; state agency responsible for assessment or certification	The student's own teacher	Moderation of marking; external checking of a sample of student products by a competent body; availability of national guidance materials for marking student performance on the examination (performance criteria, rubrics, exemplars)	Multiple choice; closed format short-answer questions; open-ended writing tasks	Performing tasks	No	National curriculum goals
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Finland	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
France	All programmes	Yes	*Diplôme national du brevet (DNB)*[4]	Yes, for all schools	All students: M, L, S, SS, FL, T, A, R	Year 9	Central education authority	Continuous assessment: the student's own teacher. For the standardised part of the examination: a teacher from another school	None	Open-ended writing tasks (essays); continuous assessment	Open-ended writing tasks (essays); continuous assessment	No	National curriculum goals
Hungary	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Iceland	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Ireland	All programmes	Yes	Junior Certificate	Yes, for public and government-dependent private schools only	All students: M, L, S, SS Students choose: A, R, V, O	Year 11	Central agency responsible for assessment (State Examination Commission)	Central agency responsible for assessment (State Examination Commission)	a	Open-ended writing tasks; closed-format short-answer questions	Written tasks (Closed-format tasks; open-ended tasks; general problem solving; problem-solving in unfamiliar contexts)	No	National curriculum goals
Israel	All programmes	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Italy	General	Yes	*Esame di Stato conclusivo del primo ciclo di istruzione (Prova nazionale)*	Yes, for all schools	All students: M, L, S, FL, A	Year 8	Central education authority (INVALSI) for the standardised central examination, which is a portion of the national 1st cycle examination	School (School Examination Committee) on the basis of INVALSI correction grids; Central agency responsible for assessment (National Institute for Assessment [INVALSI])[5]	a[5]	Multiple choice; closed-format short-answer questions; open-ended writing tasks. School-based part of the examination: written essay; oral questions and answers; oral presentation	Multiple choice; closed-format short-answer questions; open-ended writing tasks. School-based part of the examination: written problems/exercises; oral questions and answers; oral presentation	No	National curriculum goals
Korea	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Luxembourg	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Mexico	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Netherlands	General	Yes	Examination in vmbo-t programmes	Yes, for all schools	All students: L, FL, SS Students choose: M, S, FL, O	Year 12	Central agency (Central Institute for Test Development [CITO])[6]	The student's own teacher	Availability of national guidance materials; moderation of marking (systematic external moderation by school organising bodies [competent authorities])[7]	Multiple choice; open-ended writing tasks	Open-ended calculations	No	National curriculum goals
	Pre-voc and voc	Yes	Examination in vmbo-b, vmbo-k, vmbo-gt programmes	Yes, for all schools	All students: L, FL, V, SS Students choose: M, S, FL, O	Year 12	Central agency (Central Institute for Test Development [CITO])[6]	The student's own teacher	Availability of national guidance materials; moderation of marking (systematic external moderation by school organising bodies [competent authorities])[7]	Multiple choice; open-ended writing tasks	Open-ended calculations	Yes	National curriculum goals
New Zealand	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Norway	All programmes	Yes	*Sentralt gitt eksamen* (National exam)	Yes, for all schools	Sample: M, L, FL[8]	Year 10	Central education authority	Central education authority	a	Open-ended writing tasks	Closed-format short-answer questions; open-ended writing tasks	Yes, computer-based uniform technology (in mathematics only for part two of the exam)	National curriculum goals[9]

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Poland	General	Yes[10]	m	Yes, for all schools	All students: M, L, S, SS, FL	Year 9	Central and regional agencies responsible for external assessment (Central Examination Commission, Regional Examination Commissions)	Central and regional agencies responsible for external assessment (Central Examination Commission, Regional Examination Commissions)	a	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	National examination standards based on core curriculum[11]
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Portugal	General	Yes	Final national examinations	Yes, for all schools	All students: M, L	Year 9	Central education authority	Teacher from another school	Availability of national guidance materials (performance criteria); moderation of marking (teachers discussing student performance in groups)	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks; problem-solving tasks	No	National curriculum (performance criteria)
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Slovak Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Slovenia	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Spain	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Sweden	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	No	a	a	a	a	a	a	a	a	a	a	a

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Australia: There are no standardised central examinations at ISCED level 2 in Australia. However, most schools have some sort of mandatory assessments in Year 10 level. These assessments largely do not have an effect upon students' progression to a higher level of education or completion of an officially recognised degree.

2. Belgium (Fr.): Schools organise themselves for the marking of standardised central examinations. This may take different forms, e.g. correction by the teacher, another teacher, in groups, etc.

3. Denmark: 95% of government-dependent private schools also administer standardised central examinations, although they are not required to do so.

4. France: The DNB is based 60% on continuous assessment by teachers and 40% on a standardised examination marked by teachers from other schools.

5. Italy: The National Institute for Assessment (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI]) marks a standardised test worth 1/7 of the mark (which is an average of mark in national Invalsi examination, marks in non national assessments, a mark in oral colloquium and a mark in admission to the examination). There is a school-based, non standardised part of the examination which is scored locally. No regulations exist to date to moderate student marks for the non standardised part of the examination.

6. Netherlands: *Centraal Instituut voor Toetsontwikkeling*.

7. Netherlands: In a first step, marking guidelines set by a central agency, the Central Committee for Examinations, are used by the student's own teachers. In a second step, a second examiner, a teacher from another school, examines student examinations. In case of disagreement, the organising body (competent authority) of the second external examiner notifies the organising body (competent authority) of the first examiner (the student's own teacher). In case the disagreement cannot be resolved, the Inspectorate intervenes. The Inspectorate can appoint a third independent examiner that takes the ultimate decision about a student's mark.

8. Norway: All students in Year 10 sit a centrally given written examination in one subject (Norwegian, mathematics or English) and a locally given oral examination in one subject. Students are sampled randomly for the different subjects.

9. Norway: As specified in the curricula, the education law and specific subject assessment guidelines for examinations.

10. Poland: Standardised examinations at ISCED level 2 cover mathematics, science, the language of instruction, modern foreign languages and social studies. Since 2012, scores are reported separately for each subject. Between 2002 and 2011 scores were combined for mathematics and science and for the language of instruction and social studies. Scores for modern languages were already reported separately.

11. Poland: From 2012 onwards, national examination standards have been replaced by the new core curriculum, which is formulated in terms of learning outcomes.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central assessments

Belgium (Fr.): Standardised central assessments (*Évaluation externe des acquis des élèves de l'enseignement obligatoire*) are conducted for the purpose of identifying individual learning needs. They are compulsory for all students in the given years. Subjects are rotated on the basis of 3-year cycles (mathematics was tested in 2011/12).

Norway: Standardised central assessments examine competencies in reading literacy and numeracy. There are no tests in the language of instruction, modern foreign languages or mathematics.

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012)

This table describes central/state policy frameworks for standardised central assessments, i.e. standardised student assessments which are mostly used to monitor the quality of education at the system and/or school level, at ISCED level 2.

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Australia	General	Yes	National Assessment Program – Literacy and Numeracy (NAPLAN); National Assessment Program – Civics and Citizenship (NAP-CC); National Assessment Program – Information and Communication Technology Literacy (NAP-ICTL)	Yes, for all schools	All students: M (Years 7, 9); L (Years 7, 9) Sample: SS (Year 10 only); T (Year 10 only)	Years 7, 9, 10	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions	No, except computer based uniform technology is used for students with special needs and for NAP-ICTL	National learning progressions
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Austria	All programmes	Yes	*Bildungsstandards* (Educational Standards)	Yes, for public and government-dependent private schools only	All students: M, L, FL	Year 8	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[1]	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[1]	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation; oral questions and answers[2]	Multiple choice; closed-format short-answer questions; open-ended writing tasks[2]	No	National standards

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Belgium (Fl.)	All programmes	Yes	National Assessment Programme	No	Sample: varies[3]	Varies[3]	Research institute commissioned by central education authority	Research institute commissioned by central education authority	Multiple choice; closed-format short-answer questions; performing a task/experiment	Multiple choice; closed-format short-answer questions; performing a task/experiment	Yes, computer-based uniform technology for some tests	National standards
Belgium (Fr.)	General	Yes	*Évaluation externe non certificative*	Yes, for all schools	All students: varies[4]	Varies[4]	Central education authority	School	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task[4]	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task/ experiment	No	National standards[4]
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Canada	All programmes	Yes	Pan-Canadian Assessment Program (PCAP)	Yes, for public and government-dependent private schools only	Sample: M, L, S[5]	Year 8	Central education council (Council of Ministers of Education, Canada)	Central education council (Council of Ministers of Education, Canada)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	Pan-Canadian Assessment Framework
Chile	All programmes	Yes	System for Measuring the Quality of Education (SIMCE)[6]	Yes, for all schools	All students: M, L, S, SS (Year 8, every 2 years) Sample: O (physical education) (Year 8)	Year 8	Central education authority (Ministry of Education)[7]	Central education authority (Ministry of Education)[7]	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	National curriculum goals; national standards
Czech Republic	All programmes	No, but central assessments are currently being piloted[8]	Central assessment	Yes, for all schools	All students: M, L, FL	Year 9	Czech School Inspectorate (CSI)[9]	Czech School Inspectorate (CSI)[9]	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National curriculum goals; national standards

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Denmark	All programmes	Yes	*Nationale test*	No	All students: M (Year 6), L (Years 6, 8), FL (Year 7 only), S, SS (Year 8)	Varies by subject	Central agency responsible for assessment (National Agency for Quality and Supervision)	Central agency responsible for assessment (National Agency for Quality and Supervision)	Multiple choice; closed-format writing tasks; matching items (e.g. pictures/drawings with words)	Multiple choice; closed-format writing tasks; matching items	Yes, computer-based adaptive technology[10]	National learning progressions
Estonia	All programmes	No	a	a	a	a	a	a	a	a	a	a
Finland	All programmes	Yes[11]	Sample based assessments of learning outcomes	No	Sample: Subjects vary, but typically include M, L; other subjects assessed include S, FL, A, R, and cross-curricular learning-to-learn skills[9]	Varies (most regularly Year 9, also Year 7)	Central education authority (Finnish National Board of Education)	The student´s own teacher (according to central marking guidelines)	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation	Multiple choice; closed-format short-answer questions	No	Marking guidelines
France	All programmes	Yes	*Évaluations-bilan CEDRE (cycle des évaluations disciplinaires réalisées sur échantillon)* *LOLF-base (indicateurs liés à la loi organique relative à la loi des finances)*	Yes, for all schools	*Évaluations-bilan CEDRE*: Sample: A different discipline each year (6-year cycle) *LOLF-base*: Sample: M, L	Year 9	Central education authority (General Directorate for School Education [DGESCO]);[12] central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[13]	Central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[13]	Multiple choice; short open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National curriculum goals; national standards
Hungary	All programmes	Yes	National Assessment of Basic Competencies	Yes, for all schools	All students: M, L	Years 6, 8	Central education authority	Central education authority	Multiple choice; short open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National Assessment Framework
Iceland	All programmes	Yes	m	Yes, for all schools	All students: M, L, S, FL	Year 10	Central agency responsible for assessment (Educational Testing Institute)	Central agency responsible for assessment (Educational Testing Institute)	Multiple choice; open-ended writing tasks; oral questions and answers	Multiple choice; open-ended writing tasks; oral questions and answers	No	National curriculum goals
Ireland	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Israel	All programmes	Yes	*Meitzav*	Yes, for all schools according to cycles	All students: M, L, S, FL	Year 8	Central education authority or government	Central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals
Italy	General	Yes	*Servizio Nazionale di Valutazione (SNV)*	Yes, for all schools	All students: M, L	Years 6, 8	Central education authority with support of central agency responsible for assessment (National Institute for Assessment [INVALSI])[14]	Central education authority with support of central agency responsible for assessment (National Institute for Assessment [INVALSI])[14]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals[14]
Korea	All programmes	Yes	Subject Learning Diagnostic Test; National Assessment of Educational Achievement (NAEA)	Yes, for all schools	Subject Learning Diagnostic Test: All students: M, L, S, SS NAEA: All students: M, L, S, SS, FL	Subject Learning Diagnostic Test: Years 7, 8, 9 NAEA: Years 6, 9	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Luxembourg	All programmes	Yes	*Épreuves standardisées (5ème ES / 9 EST)*	Yes, for public schools only	All students: M, L (German, French)	Year 8, 10	Central education authority; research institute (University of Luxembourg)	research institute (University of Luxembourg)[15]	Multiple choice	Multiple choice	Yes, computer-based uniform technology	National standards
Mexico	All programmes	Yes	*ENLACE Básica*	Yes, for all schools	All students: M, L; further subject on rotating annual basis (e.g. S, SS)	All students: Years 7, 8, 9 Sample: Year 9[16]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation ([INEE])[17]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[17]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice	No	National curriculum goals
Netherlands	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
New Zealand	All programmes	Yes	National Education Monitoring Project	No	Sample: 4 blocks of curriculum areas assessed alternately on a 4-year cycle: (1) S, A, O (information skills [graphs, tables, maps, charts, diagrams]); (2) L, T, A; (3) M, SS, O (information skills [library, research]); (4) L, O (health and physical education)	Year 8	Central education authority or government; Private company contracted to Central education authority or government	Central education authority or government; Private company contracted to central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Yes, computer-based uniform technology	National curriculum goals
Norway	All programmes	Yes	National tests	Yes, for all schools	All students: M, L, FL	Years 8, 9	Central education authority (Norwegian Directorate for Education and Training)	The student's own teacher; central education authority (Norwegian Directorate for Education and Training)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National learning progressions
Poland	All programmes	No	a	a	a	a	a	a	a	a	a	a
Portugal	All programmes	No	a	a	a	a	a	a	a	a	a	a
Slovak Republic	General	Yes	*Testovanie 9*	Yes, for all schools	All students: M, L	Year 9	Central agency responsible for assessment (NÚCEM)	Central agency responsible for assessment (NÚCEM)	a	Multiple choice; closed-format short-answer questions	m	National curriculum goals
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Slovenia	All programmes	Yes	National Assessment	Yes, for all schools	All students: L, M, O (subject determined by the Ministry for Education and Sport)	Year 9	Central agency responsible for assessment (National Examinations Centre)	Central agencies (National Examinations Centre; National Education Institute)	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No, computer-based technology for some students with special needs only[18]	National curriculum goals
Spain	All programmes	No[19]	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Sweden	All programmes	Yes	National tests	Yes, for all schools	All students: M, L, FL, S	Year 9	Central educational authority or government; central agency (National Agency for Education)	The student's own teacher (another teacher from within the school may collaborate; Inspectorate reviews the marking of a sample of tests)	Multiple choice; open-ended writing tasks; oral presentation	Multiple choice; closed-format short-answer questions; oral presentation; oral questions and answers	No	National knowledge requirements
United Kingdom (Northern Ireland)	All programmes	Yes	Assessments using Levels of Progression (legislation is currently in train)	Yes, for all schools (from 2012/13)	All students: Communication and Using Maths[20]	Year 10 – level 5 expected	Department of Education tasks the Council for Curriculum, Examinations and Assessment (CCEA) to develop the assessments	Department of Education tasks CCEA to develop the assessments	A range of teacher-set tasks including some tasks set by CCEA	A range of teacher set tasks including some tasks set by CCEA	Yes	Levels of Progression as set out in legislation and subsequently in guidance developed by CCEA

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Austria: *Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens.*

2. Austria: From school year 2011/12 onwards.

3. Belgium (Fl.): There is no systematic assessment of particular learning areas or educational levels. At ISCED level 2 participation in standardised central assessments varies significantly among students in different programmes of secondary education.

4. Belgium (Fr.): Each year, the central assessments cover a different subject (mathematics, language of instruction or science [*éveil-initiation scientifique*]). Subjects are rotated on the basis of a 3-year cycle (e.g. mathematics in 2011/12). In 2011/12, the language of instruction was not assessed.

5. Canada: Mathematics, the language of instruction and science are assessed on a 3-year cycle focussing on one of these three subjects.

6. Chile: *Sistema de Medición de Calidad de la Educación*.

7. Chile: The National Agency for the Quality of Education (*Agencia de Calidad*) will take over these responsibilities from the Ministry of Education in October 2012.

8. Czech Republic: Information provided for the Czech Republic refers to the standardised central assessments being piloted in the academic year 2011/12.

9. Czech Republic: *Česká školní inspekce*.

10. Denmark: Standardised central assessments and school leavers' final examination after Years 9/10 are computer-based. Marking and the calculation of test scores is performed automatically.

11. Finland: The Ministry of Education and Culture has formulated a plan for assessing learning outcomes in basic education for the years 2012-15. Subjects are assessed according to the selection and timings of this plan. At the lower secondary level, a central assessment in mathematics was organised in Year 9 in 2011 and in 2012. An assessment in the language of instruction was conducted in Year 9 in 2010.

12. France: *Direction générale de l'enseignement scolaire*.

13. France: *Direction de l'évaluation, de la prospective et de la performance*.

14. Italy: The National Institute for Assessment (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI]) has developed frameworks for central student assessments in the language of instruction and mathematics which are based on the national curriculum goals.

15. Luxembourg: The tests are automatically marked by a computer.

16. Mexico: Assessed on a 4-year cycle.

17. Mexico: *Instituto Nacional para la Evaluación de la Educación*.

18. Slovenia: From 2013 onwards, the marking of standardised central assessments at ISCED level 2 will be computer based.

19. Spain: Currently, there are no standardised central assessments in Spain. The General Diagnosis Assessment was discontinued in ISCED 2 after 2010. The Spanish government is currently preparing a new education law and a different assessment design for the General Diagnosis Assessment to be re-introduced in the future in line with the new education programme. There are also regional assessment programmes undertaken under the responsibility of the Autonomous Communities.

20. United Kingdom (Northern Ireland): From 2013/14, Using ICT will also be assessed.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Annex 4.A2 Features of student assessment frameworks in primary and upper secondary education

The tables below provide information on features of student assessment frameworks in primary and upper secondary education in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Part of the information was supplied by countries through a questionnaire specifically developed by the OECD Review. The rest of the information is based on data provided by OECD *Education at a Glance 2011*.

This annex is available on line only at *http://dx.doi.org/10.1787/9789264190658-8-en*. The online annex includes the following material:

Table 4.A2.1a Internal summative assessment frameworks at ISCED level 1 (2012)

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012)

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012)

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012)

Table 4.A2.4a Standardised central examinations at ISCED level 1 (2012)

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012)

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012)

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012)

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Annex 4.A2 Features of student assessment frameworks in primary and upper secondary education

The tables below provide information on features of student assessment frameworks in primary and upper secondary education in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Part of the information was supplied by countries through a questionnaire specifically developed by the OECD Review. The rest of the information is based on data provided by OECD *Education at a Glance 2011*.

This annex is available on line only at *http://dx.doi.org/10.1787/9789264190658-8-en*.

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

General notes

Australia: Australia is a federation of eight states and territories. There are differences in policy frameworks for evaluation and assessment across states and territories as well as between public (government) and private (non-government) schools.

Belgium (Fl., Fr.): In Belgium, education policy is the responsibility of each Community. The terms "national" and "central", therefore, refer to the highest educational authorities (Ministries of Education) of the Flemish and French Communities of Belgium.

Belgium (Fl.): For public schools, the school organising bodies are typically the central educational authority (Flemish Community) and provincial/regional and local educational authorities (provinces, cities, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Belgium (Fr.): For public schools, the school organising bodies (education networks) are typically the central educational authority (French Community) and provincial and local educational authorities (provinces, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Canada: Canada comprises ten provinces and three territories. Provincial/territorial education authorities refer to the highest level of educational authorities in Canada, as there is no federal/central department of education. There are differences in policy frameworks for evaluation and assessment across provinces and territories.

Chile: For public schools, the school organising bodies (sustainers) are typically local educational authorities (municipalities).

Czech Republic: For public schools, the school organising bodies are typically local education authorities at ISCED levels 1 and 2 and regional education authorities at ISCED level 3.

Hungary: For public schools, the school organising bodies (maintainers) are typically local and regional educational authorities. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

Ireland: School boards of management comprise members external to the school such as representatives of the patron and of the local community as well as members internal to the school such as the principal, teacher representatives and parent representatives.

Netherlands: In principle, all schools are government funded. Depending on their denomination, they can have a private (religious or pedagogy-based) or public character. For public schools, school organising bodies (competent authorities) can be local educational authorities (municipal authorities), a local governing committee with transferred powers of the municipality, or a public foundation or corporation. School organising bodies for private schools can be groups of parents, foundations or corporations.

New Zealand: School Boards of Trustees typically comprise elected members from the school community, the principal, a staff representative and a student representative (in secondary schools).

Spain: Responsibilities for education are shared between the central educational authority (Ministry of Education, Culture and Sport [*Ministerio de Educación, Cultura y Deporte*]) and state educational authorities (Regional Ministries or Departments of Education of the Autonomous Communities [*Comunidades Autónomas*]). The central educational authority executes the general guidelines of the government on education policy and regulates the basic elements or aspects of the system. The Autonomous Communities develop the central regulations and have executive and administrative competences for managing the education system in their own territory. State educational authorities refer to educational authorities at the highest level of the Autonomous Communities. Throughout the Tables, the Autonomous Communities are referred to as "state educational authorities".

United Kingdom (Northern Ireland): Following political agreement to the devolution of certain policy and legislative powers from the United Kingdom government at Westminster to a local Assembly in 1999, legislative responsibility for education in Northern Ireland was one of the functions devolved to the Assembly and to a locally elected Minister for Education. The Department of Education, Northern Ireland, provides the central governance and management of education in Northern Ireland and is referred to as the "central level" throughout the tables.

Table 4.A2.1a Internal summative assessment frameworks at ISCED level 1 (2012)

This table describes central/state policy frameworks for internal summative assessment at ISCED level 1, i.e. forms of student assessment that are not standardised, but designed and marked by students' own teachers and implemented as part of regular classroom instruction in schools. This table focuses on summative forms of internal assessment, i.e. assessment designed to provide a summary statement of student achievement at a particular point in time in order to record, mark or certify achievements that may count towards a student's year-end or final marks. Summative assessment usually occurs at the end of a learning unit, term, school year, or educational level.

Country	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
	1	2	3	4	5
Australia	Yes, at the state level for all schools	State agencies	State curriculum goals; state standards[1]	Moderation of marking; availability of state guidance materials for marking student performance on the examination	100%
Austria	Yes, at the central level for public schools and government-dependent private schools only	Central education authority or government	National curriculum goals	Competences for marking are part of initial teacher education	100%
Belgium (Fl.)	No, but certain basic requirements are set at the central level for all schools[2]	Inspectorate	National curriculum goals (attainment targets and developmental objectives)	None	m
Belgium (Fr.)	No, but certain basic requirements are set at the central level for all schools[3]	Inspectorate	National curriculum goals	Availability of guidance materials for marking student performance in the examination (provided either by education authorities or school organising bodies [education networks])	100%
Canada	No, but certain basic requirements are set at the provincial/territorial level for public schools and government-dependent private schools only	Provincial/territorial education authorities or governments and a provincial central agency in one province (Ontario)	Varies across provinces/territories	Varies across provinces/territories	0%
Chile	Yes, at the central level for all schools[4]	School (school principal and teachers)	National curriculum goals; national standards	None	100%
Czech Republic	No, but certain basic requirements are set at the central level for all schools[5]	School board or committee; Czech School Inspectorate (CSI)	National curriculum goals; national standards	None	100%

Table 4.A2.1a Internal summative assessment frameworks at ISCED level 1 (2012) (*continued*)

Country	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
	1	2	3	4	5
Denmark	No	a	a	a	a
Estonia	Yes, at the central level for all schools	School (school principal)	National curriculum goals; national standards	None	100%
Finland	No	a	a	a	a
France	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%
Hungary	Yes, at the central level for all schools	School organising bodies (maintainers) and school leaders	National curriculum standards	None	100%
Iceland	No	a	a	a	a
Ireland	Yes, at the central level for public and government-dependent private schools only	School Board of Management	National curriculum goals	None	Not specified in the policy framework
Israel	No	a	a	a	a
Italy	Yes, at the central level for all schools[6]	School board or committee (includes the school principal and all teachers)	National curriculum goals[6]	None	100%
Korea	Yes, at the central level for all schools	Respective subject committee within the school	National curriculum goals	Decision made by the School Mark Management Committee within the school	100%
Luxembourg	Yes, at the central level for public schools only	Central education authorities or government; Inspectorate	National standards	Availability of national guidance criteria (performance criteria, rubrics)	100%
Mexico	Yes, at the central level for all schools	State education authorities or governments	National curriculum goals	None	100%

Table 4.A2.1a Internal summative assessment frameworks at ISCED level 1 (2012) (*continued*)

Country	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
	1	2	3	4	5
Netherlands	Yes, at the central level for all schools	School organising bodies (competent authorities)	National curriculum goals	Moderation of marking (within schools)	100%
New Zealand	Yes, at the central level for state schools and state-integrated schools only	Central education authority (Ministry of Education); central agency (Education Review Office); School Board of Trustees	National standards	Availability of national guidance materials for marking student performance (performance criteria, exemplars); moderation of marking (within and between schools)	100%
Norway	No	a	a	a	a
Poland	Yes, at the central level for all schools[7]	School principal	National standards; curriculum defined at the school level	None	100%
Portugal	Yes, at the central level for public schools only	National education authority; pedagogical council	National curriculum goals	Moderation of marking within schools (criteria approved by the pedagogical council)	100% in Years 1, 2, 3 and 5; 75% in Year 4; 70% in Year 6[8]
Slovak Republic	No	a	a	a	a
Slovenia	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%
Spain	Yes, at the central level for all schools[9]	State inspection bodies	National curriculum goals	Availability of national guidance materials for marking student performance on the examination (performance criteria)	100%
Sweden	No	a	a	a	a
United Kingdom (Northern Ireland)	Yes, at the central level for all schools	Council for Curriculum, Examinations and Assessment (CCEA); Education and Training Inspectorate in its reviews of schools	Levels of progression moderated by the Council for Curriculum, Examinations and Assessment (CCEA)	None[10]	Not specified in the framework

Notes: a: information not applicable because the category does not apply; m: information not available.

1. Australia: A national curriculum (The Australian Curriculum) is currently being implemented and is expected to replace state-based curriculum frameworks.

2. Belgium (Fl.): At present, the Ministry of Education and Training does not write comprehensive curricula. However, the Ministry specifies attainment targets and developmental objectives (*eindtermen en ontwikkelingsdoelen*), which function as the basis for each curriculum. The attainment targets and developmental objectives are also called the "core curriculum". They are set by the Flemish central education authority. It is mandatory for schools to monitor their educational quality, which implies that each school must at least assess the extent to which students acquire the skills, knowledge and attitudes described in the attainment targets and developmental objectives. In this respect each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of its students. The Inspectorate may ask schools to present their policy. To some extent, internal summative assessment at schools is regulated by the attainment targets. Schools are also required to assess whether students have attained the objectives of the school curriculum to a satisfactory level. However, there are no regulations regarding the methods schools can use to do so.

3. Belgium (Fr.): The policy framework is provided through the work of the Inspectorate and the possibility for students to appeal summative assessment decisions made by the class council through an appeals council.

4. Chile: The policy framework sets very basic requirements, such as the number of assessments per year. It also states that every school has to define an internal protocol for summative assessment.

5. Czech Republic: The central education authority does not determine the framework for the internal summative evaluation. However, the Education Act requires every school to set its own assessment system that is approved by the school board. Internal school assessment regulations and their administration are controlled by the Czech School Inspectorate (*Česká školní inspekce* [CSI]).

6. Italy: Presidential Decree n. 122/2009 provides norms on student assessment. The national curriculum goals are not explicitly mentioned in this regulation, but they are normally considered by school committees.

7. Poland: While a policy framework exists, it is very general and leaves a great degree of autonomy to schools to define their own assessment rules. Schools are, however, required by law to define assessment regulations as part of the school statute.

8. Portugal: In Year 4, internal summative assessment constitutes 75% of students' year-end marks because the standardised central examination constitutes 25% of the year-end marks. In Year 6, internal summative assessment constitutes only 70% of students' year-end marks because the standardised central examination constitutes 30% of the year-end marks. From 2013/14 onwards the weight of the internal summative assessment will be 70% for Years 4 and 6.

9. Spain: Subject teachers are responsible for the assessment and marking of students at this level. There are, however, some constraints and guidelines. Minimum subject contents, learning objectives (both summative and formative) and assessment criteria are established at a central level. At the school level subject departments write a general programme for the year, revised by the inspection services to make sure it complies with the legal framework, establishing the assessment and marking methods and criteria. Subject departments are also responsible for making a written test (optional for those students who have previously attained the objectives of the school year), common to all school groups of the same level for year-end assessment and marking. Finally, the class teachers board has a say in the end of year assessment, especially at the end of each cycle and in relation to the attainment of the general objectives of the stage.

10. United Kingdom (Northern Ireland): Moderation by the Council for Curriculum Examinations and Assessment (CCEA) will be in place at Key Stage 1 (Year 3-4) and Key Stage 2 (Year 5-7) from 2013/14.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012)

This table describes central/state policy frameworks for internal summative assessment at ISCED level 3, i.e. forms of student assessment that are not standardised, but designed and marked by students' own teachers and implemented as part of regular classroom instruction in schools. This table focuses on summative forms of internal assessment, i.e. assessment designed to provide a summary statement of student achievement at a particular point in time in order to record, mark or certify achievements that may count towards a student's year-end or final marks. Summative assessment usually occurs at the end of a learning unit, term, school year, or educational level.

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Australia	General	Yes, at the state level for all schools	State agency	State curriculum goals; state standards[1]	Moderation of marking; availability of state guidance materials	50–100%
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)	State agency	National standards	Moderation of marking; availability of state guidance materials	100%
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	Central education authority or government	National curriculum goals	Competences for marking are included in internal teacher training	100%
Belgium (Fl.)	All programmes	No, but certain basic requirements are set at the central level for all schools[2]	Inspectorate	National curriculum goals (attainment targets and developmental objectives)	None	m
Belgium (Fr.)	All programmes	No, but certain basic requirements are set at the central level for all schools[3]	Inspectorate	National curriculum goals	Availability of general guidance materials for students' assessment (provided either by education authorities or education networks)	100%
Canada	All programmes	No, but certain basic requirements are set at the provincial/territorial level for public schools and government-dependent private schools only	Provincial/territorial education authorities or governments and a provincial agency in one province (Ontario)	Varies across provinces/territories	Varies across provinces/territories	Varies across provinces/territories

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Chile	All programmes	Yes, at the central level for all schools[4]	School (school principal and teachers)	National curriculum goals; national standards	None	100%
Czech Republic	All programmes	No, but certain basic requirements are set at the central level for all schools[5]	School board or committee; Czech School Inspectorate (CSI)	National curriculum goals; national standards	None	100%
Denmark	General	Yes, at the central level for all schools	School board	National curriculum goals	Availability of national guidance materials for marking student performance on the examination (performance criteria)	60%
	Pre-voc and voc	Yes, at the central level for all schools	School board	National curriculum goals, national standards; national learning progressions	Availability of national guidance materials for marking student performance on the examination (performance criteria)	100% / 0%[6]
Estonia	General	Yes, at the central level for all schools	School (school principal)	National curriculum goals; national standards	None	a
	Pre-voc and voc	No	a	a	a	a
Finland	General	Yes, at the central level for all schools[7]	Teachers and school leaders	National curriculum goals	None	Not specified in the policy framework
	Pre-voc and voc	Yes, at the central level for all VET providers[7]	VET providers[8]	National curriculum goals	Availability of national guidance materials for marking student performance (performance criteria described for every study model in the national framework of qualifications)[9]	Not specified in the policy framework (determined by VET providers, typically 80-100%)

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
France	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%
Hungary	All programmes	Yes, at the central level for all schools	School organising bodies (maintainers) and school leaders	National curriculum standards	None	100%
Iceland	All programmes	No[10]	a	a	a	a
Ireland	All programmes	Yes, at the central level for public and government-dependent private schools only	Teachers and school leaders	National curriculum goals	Availability of national guidance material for assessing student performance[11]	Varies depending on subject specifications
Israel	General only	Yes, at the central level for all schools	School principal	National curriculum goals	None	50%
	Pre-voc and voc	No	a	a	a	a
Italy	All programmes	Yes, at the central level for all schools[12]	School board or committee (includes the school principal and all teachers)	National curriculum goals[12]	None	100%
Korea	All programmes	Yes, at the central level for all schools	Respective subject committee within the school	National curriculum goals	Decision made by the School Mark Management Committee within the school	100%
Luxembourg	All programmes	Yes, at the central level for public schools only	Central education authorities or government; school principal	National curriculum goals	None	100%

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Mexico[13]	All programmes	Yes, at the central or state level for all public schools	Centrally managed public schools: central education authority or government Locally managed public schools: state education authorities or governments	National curriculum goals	None	100%
Netherlands	General	Yes, at the central level for all schools	School organising bodies (competent authorities)	National curriculum goals	Moderation of marking (within schools); Inspectorate monitors differences between school-based assessment and central examinations at the end of the cycle	End of year: 100%; end of cycle: 50%
	Pre-voc and voc	Yes, at the central level for all schools	Inspectorate	National curriculum goals	Moderation of marking (within schools)	100%
New Zealand	All programmes	Yes, at the central level for all schools	School Board of Trustees; central agency (New Zealand Qualifications Authority)	National standards	National guidance materials for marking student performance (performance criteria, rubrics, exemplars); moderation (within schools, systematic external moderation by external body)	Depending on the course
Norway	All programmes	No, but certain basic requirements are set at the central level for all schools[14]	Local education authorities	National curriculum goals	None	80%
Poland	All programmes	Yes, at the central level for all schools[15]	School principal	National standards; curriculum defined at the school level	None	100%
Portugal	General	Yes, at the central level for public schools only	National education authority; pedagogical council	National curriculum goals	Moderation of marking within schools (criteria approved by the pedagogical council)	70%
	Pre-voc and voc	No	a	a	a	a

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Slovak Republic	General	No, but certain basic requirements are set at the central level for all schools	School principal	National educational plans; national standards	None	m
	Pre-voc and voc	No	a	a	a	a
Slovenia	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%
Spain	All programmes	Yes, at the central level for all schools[16]	State inspection bodies	National curriculum goals	Availability of national guidance materials for marking student performance on the examination (performance criteria)	100%
Sweden	All programmes	No	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	No	a	a	a	a

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes; VET: vocational education and training.

1. Australia: A national curriculum (The Australian Curriculum) is currently being implemented and is expected to replace state-based curriculum frameworks.

2. Belgium (Fl.): At present, the Ministry of Education and Training does not write comprehensive curricula. However, the Ministry specifies attainment targets and developmental objectives (*eindtermen en ontwikkelingsdoelen*), which function as the basis for each curriculum. The attainment targets and developmental objectives are also called the "core curriculum". They are set by the Flemish central education authority. It is mandatory for schools to monitor their educational quality, which implies that each school must at least assess the extent to which students acquire the skills, knowledge and attitudes described in the attainment targets and developmental objectives. In this respect each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of its students. The Inspectorate may ask schools to present their policy. To some extent, internal summative assessment at schools is regulated by the attainment targets. Schools are also required to assess whether students have attained the objectives of the school curriculum to a satisfactory level. However, there are no regulations regarding the methods schools can use to do so.

3. Belgium (Fr.): The policy framework is provided through the work of the Inspectorate and the possibility for students to appeal summative assessment decisions made by the class council through an appeals council.

4. Chile: The policy framework sets very basic requirements, such as the number of assessments per year. It also states that every school has to define an internal protocol for summative assessment.

5. Czech Republic: The central education authority does not determine the framework for the internal summative evaluation. However, the Education Act requires every school to set its own assessment system that is approved by the school board. Internal school assessment regulations and their administration are controlled by the Czech School Inspectorate (*Česká školní inspekce* [CSI]).

6. Denmark: The weight of internal summative assessment in determining students' year-end marks is 100% or 0% depending on the exam.

7. Finland: To some extent, the national core curriculum functions as a policy framework that regulates internal assessments. At ISCED level 3 (general programme) there are no criteria for different marks. However, the national core curriculum requires that the courses in every subject should be assessed (usually with numerical marks 4-10). Some procedures are mentioned in the curriculum, for instance, that course assessment must be diverse and not solely be based on written tests, but also on continuous observation. In this respect, there is also a policy framework for ISCED level 3 in place.

8. Finland: There are no separate standardised central examinations. The implementation and assessment of skills demonstrations are based on the national framework of qualification. There are 52 different vocational qualifications, which are evaluated based on an annual plan prepared by the Ministry of Education. The assessments are based on criteria. Institutions and VET providers have to take part in assessments.

9. Finland: Moderation of marking between schools using peer review on a voluntary basis.

10. Iceland: It is not mandatory by law for schools to carry out summative assessments of students. However, in practice, most schools, but not all, have a period of summative assessment twice each year. This is based on tradition and on the payments to teachers for exam periods stipulated in wage deals between the central government and teacher unions.

11. Ireland: Guidance material is available from syllabus assessment guidelines, the State Examinations Commission and other sources. Syllabus and curricular guidelines, developed by the National Council for Curriculum and Assessment, provide assessment supports, as does material provided by the Professional Development Service for Teachers. The State Examinations Commission provides guidance for teachers in summative assessment through the publication of examination papers, marking schemes and chief examiners' reports in different subjects.

12. Italy: Presidential Decree n. 122/2009 provides norms on student assessment. The national curriculum goals are not explicitly mentioned in this regulation, but they are normally considered by school committees.

13. Mexico: At schools managed by autonomous agencies (public and private), policies for the internal summative assessment of students at ISCED level 3 are determined at the school level by autonomous and private institutions (e.g. universities).

14. Norway: There are regulations in the Education Act stating that students at this level shall receive summative overall achievement marks in all subjects in addition to ongoing formative assessment and formal formative feedback at least twice a year. The overall achievement marks based on teachers judgements are entered on the leaving certificate.

15. Poland: While a policy framework exists, it is very general and leaves a great degree of autonomy to schools to define their own assessment rules. Schools are, however, required by law to define assessment regulations as part of the school statute.

16. Spain: Subject teachers are responsible for the assessment and marking of students at this level. There are, however, some constraints and guidelines. Minimum subject contents, learning objectives (both summative and formative) and assessment criteria are established at a central level. At the school level subject departments write a general programme for the year, revised by the inspection services to make sure it complies with the legal framework, establishing the assessment and marking methods and criteria. Subject departments are also responsible for making a written test (optional for those students who have previously attained the objectives of the school year), common to all school groups of the same level for year-end assessment and marking. Finally, the class teachers board has a say in the end of year assessment, especially at the end of each cycle and in relation to the attainment of the general objectives of the stage.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012)

This table describes central/state policy frameworks for the reporting of students' summative results to students and parents at ISCED level 1. Summative results in this table comprise both students' results in standardised national examinations and internal summative assessments.

Country	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 1 on formal certificates of learning?
	1	2	3	4	5
Australia	Yes, at the central level for all schools; Yes, at the state level for all schools[1]	Varies nationally (on average twice a year)	Numerical marks; alphabetical marks; descriptive ratings; qualitative assessment	Varies nationally (on average twice a year)	No formal certificate of learning is awarded
Austria	Yes, at the central level for public schools and government-dependent private schools only	Twice a year	Numerical marks (1-5); qualitative assessment for students in Years 1 and 2	Twice a year	Mark in each of the concerned subjects
Belgium (Fl.)	No[2]	a	a	a	a
Belgium (Fr.)	No	a	a	a	a
Canada	Yes, at the provincial/territorial level for public schools and government-dependent private schools only	In most cases once at the end of the school year	Varies across provinces/territories	Varies across provinces/territories	No formal certificate of learning is awarded
Chile	Yes, at the central level for all schools	Typically 2-3 times a year	No requirement exists	Typically 2-3 times a year	No formal certificate of learning is awarded
Czech Republic	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Denmark	Yes, at the central level for public schools only[3]	Only in relation to standardised national tests in Years 2-4	Descriptive ratings[4]	Once a year[5]	No formal certificate of learning is awarded

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012) (*continued*)

Country	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 1 on formal certificates of learning?
	1	2	3	4	5
Estonia	Yes, at the central level for all schools	Twice a year	Schools can decide whether to provide numerical marks, alphabetical marks or descriptive ratings and a qualitative assessment	Once a year	No formal certificate of learning is awarded
Finland	Yes, at the central level for all schools[6]	Twice a year	Numerical marks (4-10) or descriptive ratings	No requirement exists	Typically numerical marks in each of the concerned subjects, but no national requirement exists for marking before Year 8
France	Yes, at the central level for all schools	At least 3 times a year	Numerical marks (0-20 or 0-10); descriptive ratings of competencies; qualitative assessment	At least three times a year; at the request of parents	Mark in each of the concerned subjects; summary of the student's overall strengths and weaknesses; progression to the next year level; grade point average
Hungary	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5)	No requirement exists[7]	Mark in each of the concerned subjects
Iceland	No, but certain basic requirements are set at the central level	No requirements exist	Qualitative assessment	Twice a year	No formal certificate of learning is awarded
Ireland	Yes, at the central level for public schools and government-dependent private schools only	At least once a year	Qualitative assessment; numerical marks (1-10) of standardised test results in English, reading and mathematics; descriptive ratings at the end of Years 4, 6 and 8	Once a year; additional meetings facilitated at the request of parents/teachers	No formal certificate of learning is awarded
Israel	Yes, at the central level for all schools[8]	Once a year	Descriptive ratings; qualitative assessment; numerical marks	Twice a year	Mark in each of the concerned subjects

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012) (*continued*)

Country	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 1 on formal certificates of learning?
	1	2	3	4	5
Italy	Yes, at the central level for all schools	2-3 times a year[9]	Numerical marks (1-10)[9]; short comment on marks and overall student performance and behaviour	No requirement exists[9]	Mark in each of the concerned subjects[9]
Korea	Yes, at the central level for all schools	4 times a year	Descriptive ratings and qualitative assessment	No requirement exists	Pass/fail information
Luxembourg	Yes, at the central level for all schools[10]	Twice a year	Numerical scale (1-6)[11]	3 times a year	Mark in each of the concerned subjects
Mexico	Yes, at the central level for all schools	More than 4 times a year	Numerical marks (5-10)	More than 4 times a year	Pass/fail information; overall mark; mark in each of the concerned subjects (provided in the *Certificado de Terminación de Estudios* at the end of Year 6)
Netherlands	Yes, at the central level for all schools	3 times a year	Numerical marks (0-10)	3 times a year	Advice for progression to next school level
New Zealand	Yes, at the central level for all schools	Twice a year	Qualitative assessment; descriptive ratings	No requirement exists	No formal certificate of learning is awarded
Norway	No, but certain basic requirements exist at the central level[12]	a	a	At least twice a year	a
Poland	Yes, at the central level for all schools	Twice a year	Qualitative assessment in Years 1-3; descriptive ratings in Years 4-6;[13] mark for behaviour/conduct	No requirement exists	Qualitative assessment in Years 1-3; descriptive ratings in Years 4-6;[13] mark for behaviour/conduct; comments on additional coursework and special achievements
Portugal	Yes, at central level, for all schools	3 times a year	Qualitative assessment in Years 1-4; numerical marks (1-5) in Year 4 in mathematics and Portuguese; numerical marks (1-5) in Years 5-6	3 times a year and/or at request of parent	A formal certificate of learning is awarded when requested; mark in each of the concerned subjects

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012) (*continued*)

Country	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 1 on formal certificates of learning?
	1	2	3	4	5
Slovak Republic	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Slovenia	Yes, at the central level for all schools	Twice a year	Qualitative assessment up to Year 3; numerical marks (1-5) in Years 4-6	Twice a year	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in each subject up to Year 3
Spain	Yes, at the central level for all schools[14]	3 times a year	Descriptive ratings	At the request of parents	Mark in each of the concerned subjects
Sweden	Yes, at the central level for all schools	Twice a year	Descriptive ratings and qualitative assessments	Twice a year	Mark in each of the concerned subjects[15]
United Kingdom (Northern Ireland)	Yes, at the central level for all schools	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[16]	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[16]

Notes: a: information not applicable because the category does not apply.

1. Australia: Australia is a federation of eight states/territories. Standardised examinations and internal summative assessment practices are set at the state/territory level and there is a national requirement to report student results in a nationally consistent way (A-E reporting).

2. Belgium (Fl.): Schools are obliged to monitor the progress of every student and report their observations to parents. Schools are not restricted by any regulations on how to implement progress monitoring or reporting. All relevant agreements on monitoring and evaluation have to be included in the school development plan and in the school regulations. At the start of the school year every parent signs the school regulations for agreement.

3. Denmark: Private schools may voluntarily decide to report students' summative results.

4. Denmark: Parents are provided with descriptive ratings (e.g. "Your child's test score was (1) considerably above the national average score; (2) above the national average score; (3) at the national average score; (4) below the national average score; (5) considerably below the national average score").

5. Denmark: Typically, formal meetings between teachers and parents/students are more formative in nature.

6. Finland: The information provided refers to the reporting of summative results through report cards/certificates. Assessment is based on the national core curriculum, which functions as a central policy framework. Standardised national examinations are only held at the end of ISCED level 3 (general programme). In addition to school year reports, students and parents or other guardians are to be given adequate and diverse assessment feedback.

7. Hungary: At the discretion of each school.

8. Israel: The information provided refers mainly to report cards at the end of the school year.

9. Italy: Marks are given according to law 169/2008. Decisions on formal summative feedback meetings with students/parents rest within schools, per law n. 122/2009. Typically meetings take place 2-3 times a year. Student report card according to law 169/2008 (for more information see *www.istruzione.it/web/istruzione/famiglie/valutazione*)

10. Luxembourg: Students/parents receive two formal reports at the end of the first and third trimester. Parents may, however, also meet with teachers at the end of the second trimester to discuss student progress.

11. Luxembourg: For details, see the following website: *www.men.public.lu/priorites/ens_fondamental/090723_bibliotheque/111201_intermediate_reports_cycle2.pdf*.

12. Norway: According to central regulations, teachers are required to give students and parents formal feedback at least twice a year. Schools are, however, free to choose how to do so.

13. Poland: 6: Excellent, 5: Very good, 4: Good, 3: Satisfactory, 2: Acceptable, 1: Unsatisfactory.

14. Spain: More complete reports are given only to students with special educational needs or when a learning problem arises. This typically involves the identification of learning difficulties and related advice (e.g. year repetition, curricular adaptation, etc.).

15. Sweden: Since autumn 2012 formal certificates of learning include a mark in each of the concerned subjects in Year 6 at the end of each semester.

16. United Kingdom (Northern Ireland): From 2013/14, the levels of progression achieved in Using ICT will also be provided.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012)

This table describes central/state policy frameworks for the reporting of students' summative results to students and parents at ISCED level 3. Summative results in this table comprise both students' results in standardised national examinations and internal summative assessments.

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 3 on formal certificates of learning?
		1	2	3	4	5
Australia	General	Yes, at the central level for all schools Yes, at the state level for all schools[1]	Twice a year (average only – varies nationally)	Alphabetical marks; descriptive ratings; qualitative assessment	Twice a year (average only – varies nationally)	Mark in each of the concerned subjects; tertiary admission rank
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)[1, 2]	Twice a year	Descriptive ratings; alphabetical mark	Twice a year	Statement of attainment of competencies
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	At least once a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Belgium (Fl.)	All programmes	No[3]	a	a	a	a
Belgium (Fr.)	All programmes	No	a	a	a	a
Canada	All programmes	Yes, at the provincial/territorial level, for public schools and government-dependent private schools only	In most cases once at the end of the school year	Varies across provinces/territories	Varies across provinces/territories	Varies by province/territory, year level and subject area
Chile	All programmes	Yes, at the central level for all schools	Typically 2-3 times a year	No requirement exists	Typically 2-3 times a year	Secondary Certificate and marks of every year at ISCED levels 2 and 3 for every subject
	Pre-voc and voc	Yes, at the central level for all schools	Typically 2-3 times a year	No requirement exists	Typically 2-3 times a year	Secondary Certificate and marks of every year at ISCED levels 2 and 3 for every subject plus a professional/technical diploma if the requirements (job practice and report) are fulfilled

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 3 on formal certificates of learning?
		1	2	3	4	5
Czech Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	Marks in each of the subjects covered in the leaving examination (language of instruction and up to 4 other subjects)
Denmark	General	Yes, at central level for all schools	Twice a year (or more)	Numerical marks (-2 to 12)	No requirement exists	Overall mark; mark in each of the concerned subjects
	Pre-voc and voc	Yes, at central level for all schools	No requirement exists	Numerical marks (-2 to 12); descriptive ratings	No requirement exists	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in some subjects[4]
Estonia	General	Yes, at the central level for all schools	After each subject unit (35 teaching hours)	Numerical or alphabetical marks	Once a year	Final mark in each subject forms on the basis of the course grades obtained during stage of study
	Pre-voc and voc	No	a	a	a	a
Finland	General	Yes, at the central level for all schools[5]	5-6 times a year (periodically depending on how many periods the school year consists of in each school)	Numerical marks (4-10)	No requirement exists	Overall mark; mark in each of the concerned subjects
	Pre-voc and voc	Yes, at the central level for all schools[5]	Number of reports depends of the structure of the qualification[6]	Numerical marks (1-3)	No requirement exists	Overall mark; mark in each of the concerned subjects
France	All programmes	Yes, at the central level for all schools	At least 3 times a year	Numerical marks (0-20); qualitative assessment	At least three times a year; at the request of parents	Mark in each of the concerned subjects; summary of the student's overall strengths and weaknesses; progression to the next year level; grade point average
Hungary	All programmes	Yes, at the central level for public schools only	Twice a year	Numerical marks (1-5)	No requirement exists[7]	Mark in each of the concerned subjects

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 3 on formal certificates of learning?
		1	2	3	4	5
Iceland	All programmes	No[8]	a	a	a	a
Ireland	General	No, but advice and guidance are provided by a central agency (National Council for Curriculum and Assessment [NCCA])[9]	Typically twice a year	Typically numerical/ alphabetical marks and descriptive ratings	Typically once a year; additional meetings facilitated at the request of parents/teachers	Typically alphabetical mark with numerical sub grades (e.g. A1, B3) in each of the concerned subjects.
	Pre-voc and voc	No	a	a	a	a
Israel	All programmes	Yes, at the central level for all schools[10]	Twice a year	Descriptive ratings; qualitative assessment; numerical marks	Twice a year	Mark in each of the concerned subjects
Italy	All programmes	Yes, at the central level for all schools[11]	2-3 times a year	Numerical marks (1-10)	No requirement exists[12]	Mark in each of the concerned subjects
Korea	All programmes	Yes, at the central level for all schools	4 times a year	Numerical marks (0-100) and descriptive ratings (1-9)	No requirement exists	Pass/fail information
Luxembourg	All programmes	Yes, at the central level for all schools	3 times a year	Numerical marks (0-60 points); descriptive ratings[13]	At the request of students/parents only[14]	Mark in each of the concerned subjects
Mexico[15]	All programmes	Centrally managed public schools: central level Locally managed public schools: state level	Central level: more than 4 times a year State level: m	Central level: numerical marks (5-10) State level: m	No requirement exists	Pass/fail information; overall mark; mark in each of the concerned subjects (provided in the *Certificado de Terminación de Estudios*)
Netherlands	General	Yes, at the central level for all schools	3 times a year	Numerical marks (0-10)	3 times a year	Pass/fail information; mark in each of the concerned subjects

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 3 on formal certificates of learning?
		1	2	3	4	5
New Zealand	All programmes	Yes, at the central level for all schools	Once a year	Descriptive ratings	No requirement exists	Mark (descriptive ratings) in each of the concerned subjects
Norway	All programmes	Yes, at the central level for all schools	Once a year	Numerical mark (1-6)	Twice a year	Mark in each of the concerned subjects
Poland	All programmes	Yes, at the central level for all schools	Twice a year	Descriptive ratings[16]; mark for behaviour/conduct	No requirement exists	Descriptive ratings[16]; mark for behaviour/conduct; comments on additional coursework and special achievements
Portugal	General	Yes, at central level, for all schools	3 times a year	Numerical marks (0-20)	3 times a year and/or at request of parent	A formal certificate of learning is awarded
	Pre-voc and voc	No	a	a	a	a
Slovak Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	Marks in each of the subjects covered in the leaving examination (language of instruction and up to 4 other subjects)
Slovenia	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Spain	All programmes	Yes, at the central level for all schools[17]	3 times a year	Numerical marks (0-10)	At the request of parents	Mark in each of the concerned subjects
Sweden	All programmes	Yes, at the central level for all schools	Twice a year	Alphabetical marks (A-F)	Twice a year	Mark in each of the concerned courses
United Kingdom (Northern Ireland)	All programmes	Yes, at the central level for all schools	Once a year (students receive their results individually from Awarding Organisations and receive an end of year report from their teachers)	m	Once a year	Overall mark achieved at Key Stage 4 (Years 11-12) and Key Stage 5 (Years 13-14)

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

1. Australia is a federation of eight states/territories. Standardised examinations and internal summative assessment practices are set at the state/territory level and there is a national requirement to report student results in a nationally consistent way (A-E reporting).

2. Australia: Vocational education and training (VET) qualifications approved under the Australian Qualifications Framework (AQF) at ISCED levels 2 and 3 are typically offered by the VET sector. Students may take VET qualifications as part of a general programme (e.g. through the programme VET in Schools). VET (AQF) qualifications are delivered by a Registered Training Organisation (RTO) or a school in partnership with a RTO. All schools (public and private) are able to undertake partnerships with RTOs to provide students with an opportunity to undertake VET courses.

3. Belgium (Fl.): In secondary education, the "deliberative class committee" decides whether or not students pass or fail. The deliberative class committee's decisions rely on concrete data from the file of the student. The "counselling class committee" has a formative purpose. The counselling class committee is obliged to monitor the students' performance and assess their progress. The information in primary and secondary education can be provided as a qualitative assessment or by numerical marks (overall mark and marks/subject), depending on the school policy. Students having completed their upper secondary education programme successfully receive a diploma at the end of the programme, but schools are autonomous in designing the assessment of students. Students in vocational education receive a certificate at the end of ISCED level 3 unless they complete a seventh year of upper secondary education, in which case they also receive a diploma.

4. Denmark: Qualitative assessment is provided in subjects where student performance is not scored with marks, e.g. physical education.

5. Finland: The information provided for Finland refers to the reporting of summative results through report cards/certificates. Assessment is based on the national core curriculum. Standardised national examinations are only held at the end of ISCED level 3 (general programme).

6. Finland: Skills demonstration after every study module.

7. Hungary: At the discretion of each school.

8. Iceland: Legislation or the curriculum do not specify requirements about the reporting of summative results. Typically, students/parents receive summative feedback twice a year through numerical marks (0-10).

9. Ireland: Advice and guidance are provided by the National Council for Curriculum and Assessment (NCCA), but not in form of a statute.

10. Israel: The information provided refers mainly to a report card at the end of the school year.

11. Italy: Concerning pre-vocational and vocational programmes: Legislative Decree n. 167 (14 September 2011, *Testo Unico sull'apprendistato*) indicates that from the age of 15 students can apply for apprenticeship contracts and, thereby, comply with the requirements for compulsory schooling.

12. Italy: Decisions on formal summative feedback meetings with students/parents rest within schools, per law n. 122/2009. Typically meetings take place 2-3 times a year.

13. Luxembourg: For details, see the following website: *www.men.public.lu/priorites/ens_fondamental/090723_bibliotheque/111201_intermediate_reports_cycle2.pdf*.

14. Luxembourg: Every secondary school is required to organise an information meeting once a year (first or second trimester). However, there is no obligation to provide summative feedback as part of these meetings.

15. Mexico: At schools managed by autonomous agencies (public and private), policies for the reporting of student results at ISCED level 3 are determined at the school level by autonomous and private institutions (e.g. universities).

16. Poland: 6: Excellent, 5: Very good, 4: Good, 3: Satisfactory, 2: Acceptable, 1: Unsatisfactory.

17. Spain: More complete reports are given only to students with special educational needs or when a learning problem arises. This typically involves the identification of learning difficulties and related advice (e.g. Year repetition, curricular adaptation, etc.).

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central examinations

Austria: The introduction of national examinations at ISCED level 3 starts in school year 2014/15 for general programmes, and in school year 2015/16 for pre-vocational and vocational programmes. A central agency, the Federal Institute for Education Research, Innovation and Development of the Austrian School System (*Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens* [BIFIE]) is responsible for the development of the central examination. The following subjects will be examined in a standardised form at ISCED level 3: the language of instruction (German, Croatian, Hungarian, Slovenian), foreign languages (English, French, Spanish, Italian, Latin, Greek) and mathematics. Schools/teachers will be responsible for marking the new national examination from 2014-15 onwards following guidelines developed by BIFIE.

Belgium (Fr.): Standardised central examinations at ISCED levels 2 and 3 are not mandatory. Schools decide about their administration. At ISCED level 2, 55.6% of students participated in these examinations in 2011-12. At ISCED level 3, standardised national examinations in general programmes examine one competency area in history, and standardised examinations in pre-vocational and vocational programmes examine one competency area (reading) in the language of instruction (French). Schools decide about the weight given to standardised examination results in relation to other results of the student. In 2011-12, 42.9% of students participated in these examinations.

Iceland: With the introduction of the Compulsory School Act 2008, standardised national examinations were changed to standardised national assessments.

New Zealand: Students can participate in national qualifications examinations at any stage and it is not uncommon for some students to participate at ISCED level 2. However, the majority of the students undertake the national qualifications at ISCED level 3.

Table 4.A2.4a Standardised central examinations at ISCED level 1 (2012)

This table describes central/state policy frameworks for standardised central examinations, i.e. standardised student assessments that have a formal consequence for students (e.g. impact upon a student's eligibility to progress to a higher level of education, part of a process to certify learning) for ISCED level 1.

Country	Do central examinations exist at ISCED level 1?	Name of the standardised central examination at ISCED level 1	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11	12
Australia	No	a	a	a	a	a	a	a	a	a	a	a
Austria	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fr.)	Yes	*Certificat d'études de base (CEB)*	Yes, for all schools	All students: M, L, S, SS	Year 6	Central education authority	Teachers collectively[1]	Availability of national guidance materials for marking student performance on the examination	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; performing a task/experiment	No	National standards
Canada	Yes, at the provincial level	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province
Chile	No	a	a	a	a	a	a	a	a	a	a	a
Czech Republic	No	a	a	a	a	a	a	a	a	a	a	a
Denmark	No	a	a	a	a	a	a	a	a	a	a	a
Estonia	No	a	a	a	a	a	a	a	a	a	a	a
Finland	No	a	a	a	a	a	a	a	a	a	a	a
France	No	a	a	a	a	a	a	a	a	a	a	a
Hungary	No	a	a	a	a	a	a	a	a	a	a	a
Iceland	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A2.4a Standardised central examinations at ISCED level 1 (2012) (*continued*)

Country	Do central examinations exist at ISCED level 1?	Name of the standardised central examination at ISCED level 1	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11	12
Ireland	No	a	a	a	a	a	a	a	a	a	a	a
Israel	No	a	a	a	a	a	a	a	a	a	a	a
Italy	No	a	a	a	a	a	a	a	a	a	a	a
Korea	No	a	a	a	a	a	a	a	a	a	a	a
Luxembourg	No	a	a	a	a	a	a	a	a	a	a	a
Mexico	No	a	a	a	a	a	a	a	a	a	a	a
Netherlands	No	a	a	a	a	a	a	a	a	a	a	a
New Zealand	No	a	a	a	a	a	a	a	a	a	a	a
Norway	No	a	a	a	a	a	a	a	a	a	a	a
Poland	No[2]	a	a	a	a	a	a	a	a	a	a	a
Portugal	Yes	*Provas Finais Nacionais*	Yes, for all schools	All students: M; L	Years 4 and 6	Central education authority	Teacher from another school	Availability of national guidance materials (performance criteria); moderation of marking (teachers discussing student performance in groups)	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks; problem-solving tasks	No	National curriculum (performance criteria)
Slovak Republic	No	a	a	a	a	a	a	a	a	a	a	a
Slovenia	No	a	a	a	a	a	a	a	a	a	a	a
Spain	No	a	a	a	a	a	a	a	a	a	a	a
Sweden	No	a	a	a	a	a	a	a	a	a	a	a
United Kingdom (Northern Ireland)	No	a	a	a	a	a	a	a	a	a	a	a

Notes: a: information not applicable because the category does not apply.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Belgium (Fr.): The marking process is undertaken collectively by teachers within a specific geographical area. Teachers distribute the questions to be marked among them, with each teacher marking one or several questions for each student (both their own students and students taught by other teachers).

2. Poland: Standardised student assessments at ISCED level 1 have been classified as central assessments (see Table 4.A.5.1). While scores do not have formal consequences for students and students cannot fail the assessment, participation in the central assessment is compulsory. Failure to take the assessment requires students to repeat the year and lower secondary schools may use test scores for admission of students from another school district. The standardised central assessment at ISCED level 1 covers mathematics, science and the language of instruction. However, only one score is reported for each student and the test has been constructed to cover cross-curricular skills (reading, writing, reasoning, use of information and use of information in practice).

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012)

This table describes central/state policy frameworks for standardised central examinations, i.e. standardised student assessments that have a formal consequence for students (e.g. impact upon a student's eligibility to progress to a higher level of education, part of a process to certify learning) for ISCED level 3.

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Australia	All programmes	Yes, at the state level[1]	a	No[2]	All students: L Students choose: M, S, SS, FL, T, A, R, V	Year 12	State education authorities or governments; state agency responsible for assessment or certification	State education authorities or governments; state agency responsible for assessment or certification	a	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	State curriculum goals; state standards
Austria	All programmes	No[3]	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fr.)	General	Yes[4]	*Test d'enseignement secondaire supérieur (TESS)*	No	All students in participating schools: SS	Year 12	Central education authority	School[5]	Availability of national guidance materials for marking student performance on the examination	a	a	a	None
	Pre-voc and voc	Yes[4]	*Test d'enseignement secondaire supérieur (TESS)*	No	All students in participating schools: L	Year 12 or 13 depending on type of qualification	Central education authority	School[5]	Availability of national guidance materials for marking student performance on the examination	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment	a	No	National standards
Canada	All programmes	Yes, at the provincial level	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province
Chile	All programmes	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Czech Republic	All programmes	Yes[6]	Maturita	Yes, for all schools	All students: M or FL, L Students choose: M, S, SS, FL, T, A, R, O	Year 13	Central education authority or government (Ministry of Education); special institute of the Ministry of Education (Centre for Assessment of Educational Results)	Special institute of the Ministry of Education [Centre for Assessment of Educational Results]; A teacher from another school (trained by the Centre for Assessment of Educational Results)	Teachers at individual schools are trained by the Special Institute of the Ministry of Education	Multiple choice; open-ended writing tasks; oral presentation	Multiple choice; open-ended written tasks	No[7]	Catalogue of requirements for student achievement at two levels ("basic" and "advanced") for the standardised part of the examination
Denmark	General	Yes	a	Yes, for public schools only	All students: M, L Sample: S, SS, FL, T, A, R, O	Years 10, 11, 12	Central education authority	Centrally appointed external examiners	Availability of national guidance materials for marking student performance on the examination; moderation of marking (centrally appointed external examiners)	Open-ended writing tasks; oral presentation; oral questions and answers; open-ended oral tasks (analysis)	Closed-format short-answer questions; open-ended writing tasks; oral presentation; oral questions and answers	No[8]	Performance criteria
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Estonia	General	Yes	*Riigieksam*	Yes, for public and government-dependent private schools only	All students: L (Estonian or Russian, depending on the language of instruction) Students choose: M, L (Estonian or Russian), SS, FL	Year 12	State education authorities or governments; state agency responsible for assessment or certification	State agency responsible for assessment and certification	a	Closed-format short answer tasks; open-ended writing task	Performing a task (multiple step assignment)	No	National curriculum goals
	Pre-voc and voc	No[9]	a	a	a	a	a	a	a	a	a	a	a
Finland	General	Yes	Matriculation examination test	Yes, for all schools	Students choose: M, L, S, SS, FL, R	Year 12	Central agency (Matriculation Examination Board)	Central agency (Matriculation Examination Board)	a	Open-ended writing tasks	Closed-format answer questions	No	The Matriculation Examination Board issues national guidelines on the assessment of the tests
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
France	General	Yes	*Baccalauréats*	Yes, for all schools	All students: M, L, S, SS, FL, T, R Students choose: A, V, O	Years 11, 12	Central education authority	A teacher from another school	None	Open-ended writing tasks (essays); oral questions and answers	Open-ended writing tasks (essays)	No	National curriculum goals
	Pre-voc and voc	Yes	*Brevet d'études professionnelles (BEP); Certificat d'aptitude professionnelle (CAP)*	Yes, for all schools	All students: M, L, S, SS, FL, T, R Students choose: A, V, O	Year 11	Central education authority	A teacher from another school	None	Open-ended writing tasks (essays)	Multiple choice; open-ended writing tasks (essays)	No	National curriculum goals
Hungary	General	Yes	*Érettségi vizsga* (School leaving exam)	Yes, for all schools	All students: M, L, SS, FL Students choose: S, T, A, R, V	Years 12, 13	Central education authority	Central education authority; school[10]	External chair of examination board	Open-ended writing tasks; oral presentation	Open-ended writing tasks; oral presentation	No	National examination requirements
	Pre-voc and voc	Yes[11]	*Érettségi vizsga* (School leaving exam)	Yes, for all schools	V	Years 11, 12, 13, 14	Central education authorities	Examination board	External chair and members of examination board	Open-ended writing tasks; oral presentation	Open-ended writing tasks; oral presentation	No	National examination requirements
Iceland	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Ireland	All programmes	Yes	Leaving Certificate	Yes, for public and government-dependent private schools only	All students: M, L Students choose: S, SS, FL, T, A, R, V, O	Year 14	Central agency responsible for assessment (State Examination Commission)	Central agency responsible for assessment (State Examination Commission)	a	Open-ended writing tasks; closed-format short-answer questions	Written tasks (closed-format tasks; open-ended tasks; general problem solving; problem-solving in unfamiliar contexts)	No	National curriculum goals
Israel	General	Yes	*Bagrut* (Matriculation)	Yes, for public schools only	Students choose: M, L, S, SS, FL, A, V	Years 11, 12	Central education authority	Central education authority	a	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals
	Pre-voc and voc	Yes	*Bagrut* (Matriculation)	Yes, for public schools only	Students choose: M, L (Year 12 only), S, SS, FL (Year 12 only), T, V	Years 11, 12	Central education authority	Central education authority	a	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Italy	General	Yes	*Esami di Stato conclusivi dei corsi di studio di istruzione secondaria superiore*	Yes, for all schools	All students: L Students choose: M, S, SS, FL, T, A, V	Year 13	Central education authority	School (School Examination Committee)	Part of the Examination Committee comes from different schools as nominated by regional education authorities	Open-ended writing tasks; oral questions and answers; oral presentation	Only for *Liceo Scientifico*: open-ended writing tasks; oral questions and answers; oral presentation	No	National curriculum goals
	Pre-voc and voc	Yes	*Esami di Stato conclusivi dei corsi di studio di istruzione secondaria superiore*	Yes, for all schools	All students: L Students choose: M, S, SS, FL, T, A, V	Year 14	Central education authority	School (School Examination Committee)	Part of the Examination Committee comes from different schools as nominated by regional education authorities	Open-ended writing tasks; oral questions and answers; oral presentation	According to the specific educational programme: open-ended writing tasks; oral questions and answers; oral presentation	No	National curriculum goals
Korea	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Luxembourg	All programmes	Yes	*Examens de fin d'études scolaires*	Yes, for public schools only	FL	Year 12	Central education authority	The student's own teacher; another teacher from within the school	Availability of national guidance materials for marking student performance on the examination (performance criteria)	Multiple choice; open-ended writing tasks; oral questions and answers	Multiple choice; performing a task	No	National curriculum goals
Mexico	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Netherlands	General	Yes	Examinations in havo and vwo programmes	Yes, for all schools	All vwo-students: L, FL, M, SS, A, O; all havo-students: L, FL, SS, A Students choose: (M), S, T, O	Years 13, 14	Central agency (Central Institute for Test Development [CITO])[12]	The student's own teacher	Availability of national guidance materials; moderation of marking (systematic external moderation by school organising bodies [competent authorities])[13]	Multiple choice; open-ended writing tasks	Open-ended calculations	No	National curriculum goals
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
New Zealand	All programmes	Yes	National Certificate of Educational Achievement	No	Students choose L, M, S, SS, FL, T, A, V	Years 11, 12, 13	Central education authority	School; central education authority[14]	Availability of national guidance materials for marking student performance on the examination (performance criteria, rubrics, exemplars), moderation (within schools, systematic external moderation by external body)	Time-limited examination; papers for most subjects; portfolios[15]	Scaffolded questions to solve problems in context	No[16]	National standards
Norway	General	Yes	*Sentralt gitt eksamen* (National exam)	Yes, for all schools	All students: L Sample: M, S, SS, FL, T, A, V	Years 11, 12, 13	Central education authority	Central education authority; student's own teacher; external examiner	a	Open-ended writing tasks	Closed-format short-answer questions; open-ended writing tasks	Yes, computer-based uniform technology (in mathematics only for part two of the exam)	National curriculum goals[17]
	Pre-voc and voc	Yes	*Sentralt gitt eksamen* (National exam)	Yes, for all schools	Sample: S, FL, T, A, V	Years 12, 13, 14	Central education authority	Local education authorities	a	a (Information missing)	a (Information missing)	a (Information missing)	a (Information missing)
Poland	General	Yes	m	Yes, for all schools	Students choose: M, L, S, SS, FL, T, A	Year 12	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	a	Multiple choice; open-ended writing tasks; oral presentation of student project	Multiple choice; open-ended writing tasks	No	National examination standards based on core curriculum[18]
	Pre-voc and voc	Yes	m	Yes, for all schools	Students choose: M, L, S, SS, FL, T, A, V	Years 12, 13 V also in Years 11, 14 and 15	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	a	a[19]	a[19]	a[19]	National standards based on the core curriculum for specific vocational qualifications

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the upper secondary level?	Name of the standardised central examination at ISCED level 3?	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Portugal	General	Yes	National Examinations	Yes, for all schools	All students: M, L, S, SS, FL, A	Years 11, 12	Central education authority	Teacher from another school	Availability of national guidance materials (performance criteria); moderation of marking (teachers discussing student performance in groups)	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks; problem-solving tasks	No	National curriculum (performance criteria)
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Slovak Republic	All programmes	Yes	Maturita	Yes, for all schools	All students: M, L, FL	Year 13	Central agency responsible for assessment (NÚCEM)	Central agency responsible for assessment (NÚCEM)	a	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Slovenia[20]	General	Yes	General Matura	Yes, for all schools	All students: L, M, FL Students choose: 2 out of S, SS, FL, T, A	Year 13	Central agency responsible for assessment (National Examinations Centre)	Central agency responsible for assessment (National Examinations Centre)	a	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral questions and answers	No, computer-based technology only for some students with special needs	National curriculum goals
	Pre-voc and voc	Yes	Vocational Matura	Yes, for all schools	All students: L, T, V Students choose: M, FL	Years 13, 14	School; central agency responsible for assessment (National Examinations Centre)	School	Availability of national guidance materials for marking student performance on the examination	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral questions and answers	No, computer-based technology only for some students with special needs	National curriculum goals
Spain	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Sweden	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	Yes	GCSE; A-level (or equivalent)	Yes (all students would be expected to undertake some form of examination at Key Stage 4)	All students: M, L Students choose: S, SS, FL, T, A, R, V, O	GCSE: Year 12 A-level: Year 14	Schools are free to choose from a range of Awarding Organisations	Relevant Awarding Organisations (AOs) provide the external assessment. Where internal assessment is also used, teachers are involved and this is moderated by the relevant AO	Availability of AO guidance materials for marking student performance on the examination; moderation of marking (centrally appointed external examiners)	Open-ended writing tasks; closed-format short-answer questions	Written tasks (closed-format tasks; open-ended tasks; general problem solving; problem-solving in unfamiliar contexts)	To a limited extent, yes (depending upon AO)	Central curriculum goals

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Australia: All Australian states and territories have separate standardised examinations at ISCED level 3. This table presents information on the average situation for standardised examinations at the state level. Students' exam results are moderated by state-level bodies to give eligible students a higher school certificate, which provides information about how well the student has achieved in each course undertaken. This certificate may be used to apply to progress to a higher level of education. The majority of states also supply eligible students with an Australian Tertiary Admission Rank (ATAR), which is a rank determined by aggregating scaled marks from the state exam. The ATAR provides information about how well a student has performed overall against other students and is used by students to apply to progress to a higher level of education. Vocational programmes may be included in central examinations in some states or territories (for example the Victorian Certificate for Education) and, for students that undertake them, have an effect on students' progression to a higher level of education or completion of an officially-recognised competency. Tasmania, on the other hand, is an example of a state which does not include vocational programmes in their central examinations.

2. Australia: While standardised central examinations at ISCED level 3 are not compulsory to administer, 100% of public schools and 99% of government-dependent private schools do administer standardised national examinations.

3. Austria: Standardised central examinations will be introduced in 2014/15 (general programmes) and 2015/16 (Pre-vocational and vocational programmes). A central agency, the Federal Institute for Education Research, Innovation and Development of the Austrian School System (*Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens* [BIFIE]), is responsible for the development of the central examination. The following subjects will be examined in a standardised manner: language of instruction (German, Croatian, Hungarian, Slovenian), foreign languages (English, French, Spanish, Italian, Latin, Greek) and mathematics. The marking of the new central examination (from 2013-14 on) will be at the discretion of schools/teachers according to guidelines developed by BIFIE.

4. Belgium (Fr.): The standardised central examination in general programmes assesses student competencies in history, but only a part of it (*compétence critique*). The central standardised examination in pre-vocational and vocational programmes is partial and assesses French-language skills ("reading information texts"). The school decides the weight given to the results of examinations in relation to the rest of the student's performance. In 2011/12, 42.9% of students took part in this examination which is at the discretion of schools to administer. All students in participating schools take the examination.

5. Belgium (Fr.): Schools organise themselves for the marking of standardised central examinations. This may take different forms, e.g. correction by the teacher, another teacher, in groups, etc.

6. Czech Republic: The examination at ISCED level 3 has two components: (1) a standardised central component described in this table; and (2) a school-based component. The school-based component consists of two or three assessments set in Framework Education Programmes (FEPs) and assessments in two optional subjects.

7. Czech Republic: Computer-based technology is used for answer sheets scanning and subsequent data management.

8. Denmark: There are written ICT-based exam assignments with students having access to the internet.

9. Estonia: There are no compulsory central examinations in vocational schools, except for students in study-groups where Russian is the language of instruction, who have to pass an examination in Estonian as a foreign language in Year 12. However, students in vocational schools may decide to pass the same national examinations as students in general programmes. They may also pass the professional examinations which are necessary for awarding occupational qualifications.

10. Hungary: There are two types of examinations: (1) Marking is carried out at the school at the "normal level", and (2) by central assessors at the "advanced level".

11. Hungary: This applies to upper secondary vocational schools only and not to vocational training schools.

12. Netherlands: *Centraal Instituut voor Toetsontwikkeling*.

13. Netherlands: In a first step, marking guidelines set by a central agency, the Central Committee for Examinations, are used by the student's own teachers. In a second step, a second examiner, a teacher from another school, examines student examinations. In case of disagreement, the school organising body (competent authority) of the second external examiner notifies the school organising body (competent authority) of the first examiner (the student's own teacher). In case the disagreement cannot be resolved, the Inspectorate intervenes. The Inspectorate can appoint a third independent examiner that takes the ultimate decision about a student's mark.

14 New Zealand: Internally assessed subjects are marked within the school whereas externally assessed papers are marked by a central education authority (New Zealand Qualifications Authority).

15. New Zealand: Closed-format short answer questions are also rarely used.

16. New Zealand: Examinations are generally paper-based. Computer-based technology is used for the administration of examinations, i.e. inputting marks etc.

17. Norway: As specified in the curricula, the education law and specific subject assessment guidelines for examinations.

18. Poland: From 2015, national examination standards will be replaced by the new core curriculum, which is formulated in terms of learning outcomes.

19. Poland: At ISCED level 3, different forms of national standardised examinations are available for students in pre-vocational and vocational programmes. Students at ISCED level 3A programmes in vocational schools (technical schools) can choose to take the graduation examinations (maturity exam) in general subjects required for entry into higher education or a vocational qualification exam which covers practical and vocational skills. Students at ISCED level 3C schools (basic vocational schools) can sit vocational qualification exams only.

20. Slovenia: Ca. 40% of a yearly cohort ends general education at ISCED level 3 with the General Matura, ca. 45% of a yearly cohort ends vocational education at ISCED level 3 with the Vocational Matura. In addition, a further 15% of a yearly cohort ends vocational education at ISCED level 3 without direct access to tertiary education. These two and three-year programmes also entail a final examination that is compulsory after 11 or 12 years of schooling in all schools. This final examination is school-based and does not constitute a standardised central examination. Compulsory subjects include the language of instruction (L) and a practical exam (V). For the language of instruction (L) students have to complete open-ended writing tasks and an oral examination. For the practical exam (V) students have to perform practical work, explain and defend it.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central assessments

Belgium (Fr.): Standardised central assessments (*Évaluation externe des acquis des élèves de l'enseignement obligatoire*) are conducted for the purpose of identifying individual learning needs. They are compulsory for all students in the given years. Subjects are rotated on the basis of three-year cycles (mathematics was tested in 2011/12).

Norway: Standardised central assessments examine competencies in reading literacy and numeracy. There are no tests in the language of instruction, modern foreign languages or mathematics.

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012)

This table describes central/state policy frameworks for standardised central assessments, i.e. standardised student assessments which are mostly used to monitor the quality of education at the system and/or school level, at ISCED level 1.

Country	Do central assessments exist at ISCED level 1?	Name of the central assessment at ISCED level 1	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11
Australia	Yes	National Assessment Program – Literacy and Numeracy (NAPLAN); National Assessment Program – Science Literacy (NAP-SL); National Assessment Program – Civics and Citizenship (NAP-CC); National Assessment Program – Information and Communication Technology Literacy (NAP-ICTL)[1]	Yes, for all schools	All students: M, L (Years 3, 5 only) Sample: S, SS, T (Year 6 only)	NAPLAN: Years 3, 5 NAP-SL, NAP-CC, NAP-ICTL: Year 6 (triennial tests)	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions	No, except computer-based uniform technology is used for students with special needs and NAP-ICTL	National learning progressions
Austria	Yes	*Bildungsstandards* (educational standards)	Yes, for public and government-dependent private schools only	All students: M, L	Year 4	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[2]	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[2]	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation; oral questions and answers[3]	Multiple choice; closed-format short-answer questions; open-ended writing tasks[3]	No	National standards
Belgium (Fl.)	Yes	National Assessment Programme	No	Sample: varies[4]	Varies[4]	Research institute commissioned by central education authority	Research institute commissioned by central education authority	Multiple choice; closed-format short-answer questions; performing a task/experiment	Multiple choice; closed-format short-answer questions; performing a task/experiment	Yes, computer-based uniform technology for some tests	National standards

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012) (*continued*)

Country	Do central assessments exist at ISCED level 1?	Name of the central assessment at ISCED level 1	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11
Belgium (Fr.)	Yes	*Évaluation externe non certificative*	No	All students: varies[5]	Years 2, 5	Central education authority	Teachers of the school collectively[6]	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task[5]	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task/experiment	No	National standards[5]
Canada	No	a	a	a	a	a	a	a	a	a	a
Chile	Yes	System for Measuring the Quality of Education (SIMCE)[7]	Yes, for all schools	All students: L (Years 2 [reading only] and 4), M, S, SS (Year 4 only) Sample: L (writing)	All students: Years 2, 4 Sample: Year 6	Central education authority (Ministry of Education)[8]	Central education authority (Ministry of Education)[8]	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	National curriculum goals; national standards
Czech Republic	No, but central assessments are currently being piloted[9]	Central assessment	Yes, for all schools	All students: M, L, FL	Year 5	Czech School Inspectorate (CSI)[10]	Czech School Inspectorate (CSI)[10]	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National curriculum goals; national standards
Denmark	Yes	*Nationale test*	No	All students: M (Year 3 only), L (Years 2 and 4 only)	Varies by subject	Central agency responsible for assessment (National Agency for Quality and Supervision)	Central agency responsible for assessment (National Agency for Quality and Supervision)	Multiple choice; closed-format writing tasks; matching items (e.g. pictures/drawings with words)	Multiple choice; closed-format writing tasks; matching items	Yes, computer-based adaptive technology[11]	National learning progressions
Estonia	Yes	National sampling (*tasemetöö*)	Yes, for sampled schools	Sample: M, L (Years 3, 6), additional subject announced in the same year (Year 6)	Years 3, 6	Central agency responsible for assessment or certification	The student's own teacher	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals
Finland	Yes[12]	Sample based assessments of learning outcomes	No	Sample: Subjects vary, but typically include M, L; other subjects assessed include FL, S, R, A (Years 3; 5) and cross-curricular learning-to-learn skills (Years 3; 6)[12]	Varies (typically Years 3, 5, 6)	Central education authority (Finnish National Board of Education)	The student´s own teacher (according to central marking guidelines)	Closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals for the development of tasks; marking guidelines for each open-ended task

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012) (*continued*)

Country	Do central assessments exist at ISCED level 1?	Name of the central assessment at ISCED level 1	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11
France	Yes	*Évaluations nationales de CE1 et de CM2* *Évaluations-bilan CEDRE (cycle des évaluations disciplinaires réalisées sur échantillon)* *LOLF-base (indicateurs liés à la loi organique relative à la loi des finances)*	Yes, for public and government-dependent private schools only	*Évaluations nationales*: All students: M, L *Évaluations-bilan CEDRE*: Sample: 6 disciplines assessed alternately on a 6-year cycle *LOLF-base*: Sample: M, L	*Évaluations nationales*: Years 2, 5 *Évaluation-bilan CEDRE* and *LOLF-base*: Year 5	Central education authority (General Directorate for School Education [DGESCO]);[13] General Inspectorate; central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[14]	The student's own teacher; central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[14]	Open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National curriculum goals; national standards
Hungary	Yes	Diagnostic assessment	No[15]	All students: M, L	Year 4	Central education authority (OH)	Teachers	Joining, classifying, ordering, answer pool	Multiple choice; closed-format short-answer questions	No	Marking guidelines
Iceland	Yes		Yes, for all schools	All students: M, L, S	Years 4, 7	Central agency responsible for assessment (Educational Testing Institute)	Central agency responsible for assessment (Educational Testing Institute)	Multiple choice; performing a task/experiment	Multiple choice; performing a task/experiment	No	National curriculum goals
Ireland	Yes	National standardised testing[16] National assessments	National standardised testing: yes, for all schools National assessment: yes, for schools in the sample	National standardised testing: All students: M, L National assessment: Sample: M, L[17]	National standardised testing: Years 4, 6, 8 National assessment: Years 4, 8	National standardised testing: Independent assessment providers National assessment: Independent, state-funded research institution (Educational Research Centre)	National standardised testing: Schools (can also avail of external marking service) National assessment: Independent, state-funded research institution (Educational Research Centre)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Israel	Yes	*Meitzav*	Yes, for all schools according to cycles	All students: M, S, SS, FL (Year 5), L (Year 2) Sample: R (Years 4, 5)	Varies by subject	Central education authority or government (National Authority for Measurement and Evaluation in Education [RAMA])	Central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals
Italy	Yes	*Servizio Nazionale di Valutazione (SNV)*	Yes, for all schools	All students: M, L	Years 2, 5	Central agency responsible for assessment (National Institute for Assessment [INVALSI])[18]	Central agency responsible for assessment (National Institute for Assessment [INVALSI])[18]	Multiple choice; reading test (Year 2); closed-format short-answer questions (Year 5)	Multiple choice; closed-format short-answer questions; open-ended writing tasks (Year 5)	No	National curriculum goals[18]

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012) (*continued*)

Country	Do central assessments exist at ISCED level 1?	Name of the central assessment at ISCED level 1	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11
Korea	Yes	Subject Learning Diagnostic Test	Yes, for all schools	All students: M, L, S, SS, FL	Years 4, 5, 6	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Luxembourg	Yes	*Épreuves standardisées (Cycle 3.1)*	Yes, for public schools only	All students: L (German), M	Year 5	Central education authority; research institute (University of Luxembourg)	Central education authority; research institute (University of Luxembourg)	Multiple choice; performing a task/experiment; oral questions and answers	Multiple choice; performing a task/experiment	No	National standards
Mexico	Yes	*ENLACE Básica*	Yes, for all schools	All students: M, L, subject on rotating annual basis (e.g. S, SS) Sample: M, L, S, SS	Years 3, 6[19]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[20]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[20]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice	No	National curriculum goals
Netherlands	Yes	Periodic National Assessment (PPON)	No	Sample: M, L, FL, SS, S, O	Year 5 (M, L only), Year 8 (M, L, FL, SS, S, O)	Central agency (Central Institute for Test Development [CITO])[21]	Central agency (Central Institute for Test Development [CITO])[21]	Multiple choice; open-ended writing tasks	Multiple choice	No	National curriculum goals
New Zealand	Yes	National Monitoring Project	No	Sample: 4 blocks of curriculum areas assessed alternately on a 4-year cycle: (1) S, A, O (information skills [graphs, tables, maps, charts, diagrams]); (2) L, T, A; (3) M, SS, O (information skills [library, research]); (4) L, O (health and physical education)	Year 4	Central education authority or government; private company contracted to central education authority or government	Central education authority or government; private company contracted to central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Yes, computer-based uniform technology	National curriculum goals
Norway	Yes	National tests and mapping tests	Yes, for all schools	All students: M, L, FL	Year 5	Central education authority (Norwegian Directorate for Education and Training)	The student's own teacher; central education authority (Norwegian Directorate for Education and Training)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National learning progressions

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012) (*continued*)

Country	Do central assessments exist at ISCED level 1?	Name of the central assessment at ISCED level 1	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
	1	2	3	4	5	6	7	8	9	10	11
Poland	Yes	m	Yes, for all schools	All students: O (cross-curricular skills [reading, writing, reasoning, use of information])	Year 6	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	Central and regional agencies responsible for external assessment (Central Examination Commission; Regional Examination Commissions)	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	Examination standards based on core curriculum
Portugal	No	a	a	a	a	a	a	a	a	a	a
Slovak Republic	No	a	a	a	a	a	a	a	a	a	a
Slovenia	Yes[22]	National Assessment	Yes, for all schools	All students: L, M, FL	Year 6	Central agency responsible for assessment (National Examinations Centre)	School	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No, computer-based technology only for some students with special needs	National curriculum goals
Spain	No[23]	a	a	a	a	a	a	a	a	a	a
Sweden	Yes	National tests	Yes, for all schools	All students: M, L, FL (Year 6 only)	Years 3, 6[24]	Central educational authority or government; central agency (National Agency for Education)	The student's own teacher (another teacher from within the school may collaborate; Inspectorate reviews the marking of a sample of tests)	Multiple choice; closed-format short answers; open-ended writing tasks; oral presentation; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation in Year 6; oral questions and answers; performing a task/experiment	No	National knowledge requirements
United Kingdom (Northern Ireland)	Yes	Assessments using Levels of Progression (legislation is currently in train)	Yes, for all schools (from 2012/13)	All students: Communication and Using Maths[25]	Year 4 – level 2 expected Year 7 – level 4 expected	Department of Education tasks the Council for Curriculum, Examinations and Assessment (CCEA) to develop the assessments	Department of Education tasks CCEA to develop the assessments	A range of teacher-set tasks including some tasks set by CCEA	A range of teacher set tasks including some tasks set by CCEA	Yes	Levels of Progression as set out in legislation and subsequently in guidance developed by CCEA

Notes: a: information not applicable because the category does not apply; m: information not available.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test

1. Australia: In addition to the National Assessment Programme, Australian states and territories also have their own standardised central assessments. Examples include the English Online Interview for students in Years 1 and 2 (Victoria) and the Performance Indicators in Primary Schools (PIPs) for Preparatory Students in both the Australian Capital Territory and Tasmania.

2. Austria: *Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens*.

3. Austria: From school year 2012/13 onwards. Standardised central assessments for mathematics in Year 4 will be carried out for the first time in 2012-13. Standardised central assessments in the language of instruction will be carried out in Year 4 in 2013-14.

4. Belgium (Fl.): There is no systematic assessment of particular learning areas or educational levels. At ISCED level 1, for example, mathematics was assessed in 2002 and 2009; Dutch in 2002 and 2007; French in 2008; and technology in 2012.

5. Belgium (Fr.): Each year, the central assessments cover a different subject (mathematics, language of instruction or science [*éveil-initiation scientifique*]). Subjects are rotated on the basis of a three-year cycle (e.g. mathematics in 2011-12). In 2011/12, the language of instruction was not assessed.

6. Belgium (Fr.): The marking process is undertaken collectively by teachers within a specific geographical area. Teachers distribute the questions to be marked among them, with each teacher marking one or several questions for each student (both their own students and students taught by other teachers).

7. Chile: *Sistema de Medición de Calidad de la Educación*.

8. Chile: The National Agency for the Quality of Education (*Agencia de Calidad*) will take over these responsibilities from the Ministry of Education in October 2012.

9. Czech Republic: Information provided for the Czech Republic refers to the standardised central assessments being piloted in the academic year 2011/12.

10. Czech Republic: *Česká školní inspekce.*

11. Denmark: Standardised central assessments and school leavers' final examination after Years 9/10 are computer-based. Marking and the calculation of test scores is performed automatically.

12. Finland: The Ministry of Education and Culture has formulated a plan for assessing learning outcomes in basic education for the years 2012-2015. Subjects are assessed according to the selection and timings of this plan. During the school year 2011/2012, there were no sample-based national assessments at the primary level. The assessment has typically assessed the language of instruction and mathematics at the primary level. The last national assessment at the primary level was carried out in mathematics in 2008. In addition there was an assessment in science (environment and nature studies) in 2006.

13. France: *Direction générale de l'enseignement scolaire*.

14. France: *Direction de l'évaluation, de la prospective et de la performance.*

15. Hungary: Standardised central assessments were compulsory for all schools until 2011, but not from 2012 onwards. The responsible central education authority (OH) continues to publish the test materials and the software to analyse the results online.

16. Ireland: There are two sets of standardised central assessments available from independent test developers for use by schools (The Drumcondra Tests [English, Irish and mathematics] and Micra-T [English]/Sigma-T [mathematics]). The 2011 National Literacy and Numeracy Strategy requires schools to report aggregate student results in these tests to their school boards (school board of management) and to the Department of Education and Skills.

17. Ireland: National standardised testing: All schools are expected to administer standardised tests in English reading and mathematics on an annual basis. Irish-medium schools are also expected to administer tests in Irish reading. National assessment: English Reading and Mathematics (representative sample of schools) and Irish reading (Irish-medium schools only) are assessed on a 5-year cycle.

18. Italy: The National Institute for Assessment (*Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI])* has developed frameworks for central student assessments in the language of instruction and mathematics which are based on the national curriculum goals.

19. Mexico: Each year level assessed on a 4-year cycle.

20. Mexico: *Instituto Nacional para la Evaluación de la Educación*.

21. Netherlands: *Centraal Instituut voor Toetsontwikkeling*.

22. Slovenia: Schools participate on a voluntary basis.

23. Spain: Currently, there are no standardised central assessments in Spain. The General Diagnosis Assessment was discontinued in ISCED 1 after 2009. The Spanish government is currently preparing a new education law and a different assessment design for the General Diagnosis Assessment to be re-introduced in the future in line with the new education programme. There are also regional assessment programmes undertaken under the responsibility of the Autonomous Communities.

24. Sweden: This information is new for 2012.

25. United Kingdom (Northern Ireland): From 2013/14, Using ICT will also be assessed.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012)

This table describes central/state policy frameworks for standardised central assessments, i.e. standardised student assessments which are mostly used to monitor the quality of education at the system and/or school level, at ISCED level 3.

Country	Programme type	Do central assessments exist at ISCED level 3?	Name of the central assessment at ISCED level 3	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Australia	All programmes	No	a	a	a	a	a	a	a	a	a	a
Austria	All programmes	No	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	All programmes	Yes	National Assessment Programme	No	Sample: varies[1]	Varies[1]	Research institute commissioned by central education authority	Research institute commissioned by central education authority	Multiple choice; closed-format short-answer questions; performing a task/experiment	Multiple choice; closed-format short-answer questions; performing a task/experiment	Yes, computer-based uniform technology for some tests	National standards
Belgium (Fr.)	All programmes	Yes	*Évaluation externe non certificative*	Yes, for all schools	All students: varies[2]	Varies	Central education authority	School	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task/experiment	No	National standards
Canada	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 3?	Name of the central assessment at ISCED level 3	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Chile	All programmes	Yes	System for Measuring the Quality of Education (SIMCE)[3]	Yes, for all schools	All students: M, L, S, SS (Year 10); FL (Year 11)	Year 10, 11	Central education authority (Ministry of Education)[4]	Central education authority (Ministry of Education)[4]	Multiple choice; open-ended writing tasks	Multiple choice	No	National curriculum for Years 9, 10 and 11
Czech Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a
Denmark	All programmes	No	a	a	a	a	a	a	a	a	a	a
Estonia	All programmes	No	a	a	a	a	a	a	a	a	a	a
Finland	All programmes	No	a	a	a	a	a	a	a	a	a	a
France	All programmes	No	a	a	a	a	a	a	a	a	a	a
Hungary	All programmes	Yes	National Assessment of Basic Competencies	Yes, for all schools	All students: M, L	Year 10	Central education authority	Central education authority	Multiple choice; short open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National Assessment Framework
Iceland	All programmes	No	a	a	a	a	a	a	a	a	a	a
Ireland	All programmes	No	a	a	a	a	a	a	a	a	a	a
Israel	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 3?	Name of the central assessment at ISCED level 3	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Italy	All programmes	Yes	*Servizio Nazionale de Valutazione (SNV)*	Yes, for all schools	All students: M, L	Year 10	Central agency responsible for assessment (National Institute for Assessment [INVALSI])[5]	Central agency responsible for assessment (National Institute for Assessment [INVALSI])[5]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals[5]
Korea	All programmes	Yes	National Assessment of Educational Achievement (NAEA)	Yes, for all schools	All students: M, L, S, SS, FL	Year 10	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Luxembourg	All programmes	No	a	a	a	a	a	a	a	a	a	a
Mexico	All programmes	Yes	*ENLACE Media Superior*	Yes, for all schools	All students: M, L; further subject on rotating annual basis (e.g. S, SS)	Year 12[6]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[7]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[7]	Multiple choice	Multiple choice	No	National curriculum goals (language of instruction)

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 3?	Name of the central assessment at ISCED level 3	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Netherlands	All programmes	No	a	a	a	a	a	a	a	a	a	a
New Zealand	All programmes	No	a	a	a	a	a	a	a	a	a	a
Norway	All programmes	No	a	a	a	a	a	a	a	a	a	a
Poland	All programmes	No	a	a	a	a	a	a	a	a	a	a
Portugal	All programmes	No	a	a	a	a	a	a	a	a	a	a
Slovak Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a
Slovenia	All programmes	No	a	a	a	a	a	a	a	a	a	a
Spain	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 3?	Name of the central assessment at ISCED level 3	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Sweden	All programmes	Yes	National Tests for Upper Secondary Education	Yes, for all schools	All students: M, L, S, FL	Years 10, 11, 12	Central education authority or government; central agency (National Agency for Education)	The student's own teacher (another teacher from within the school may collaborate; Inspectorate reviews the marking of a sample of tests)	Multiple choice; open-ended writing tasks; oral presentation	Multiple choice; closed-format short-answer questions; oral presentation; oral questions and answers	No	National knowledge requirements
United Kingdom (Northern Ireland)	All programmes	No	a	a	a	a	a	a	a	a	a	a

Notes: a: information not applicable because the category does not apply.

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Belgium (Fl.): There is no systematic assessment of particular learning areas or educational levels. Participation in standardised central assessments varies significantly among students in different programmes of secondary education.

2. Belgium (Fr.): Each year, the central assessments cover a different subject (mathematics, language of instruction or science). Subjects are rotated on the basis of a three-year cycle (e.g. mathematics in 2011-12). In 2011/12, the language of instruction was not assessed.

3. Chile: *Sistema de Medición de Calidad de la Educación*.

4. Chile: The National Agency for the Quality of Education (*Agencia de Calidad*) will take over these responsibilities from the Ministry of Education in October 2012.

5. Italy: The National Institute for Assessment (*Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI])* has developed frameworks for central student assessments in the language of instruction and mathematics which are based on the national curriculum goals.

6. Mexico: Assessed on a 4-year cycle.

7. Mexico: *Instituto Nacional para la Evaluación de la Educación.*

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Chapter 5

Teacher appraisal:

Enhancing teacher professionalism

Improving the quality and equity of schooling depends to a large extent on the motivation and performance of individual teachers in the classroom. In turn, effective appraisal and feedback for teachers is essential to increase the focus on teaching quality and teachers' professional learning. Teacher appraisal can also support the effective organisation of schools by allowing teachers to progress in their career and take on new roles and responsibilities based on a solid evaluation of their performance. This chapter describes the approaches that countries take to appraise individual teachers. Building on a discussion of impact, drivers and contextual developments, it analyses the governance of teacher appraisal schemes, appraisal procedures and instruments, capacities required for effective appraisal and the use of appraisal results for different purposes. The chapter concludes with a set of pointers for policy development.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter looks at approaches to teacher appraisal within the evaluation and assessment framework. Teacher appraisal refers to the evaluation of teachers to make a judgement and/or provide feedback about their competencies and performance. It typically aims to support teachers' professional development and/or career advancement, and also serves to hold teachers accountable for their practice. This chapter covers different types of teacher appraisal, including appraisal for the completion of a probationary period, appraisal for performance management and reward schemes.

The effective monitoring and appraisal of teachers is central to the continuous improvement of schooling. Teachers need feedback on their performance to help them identify how to better shape and improve their teaching practice and, with the support of effective school leadership, to develop schools as professional learning communities. Teacher appraisal also provides opportunities to incentivise, recognise and reward effective teaching. Hence, the development of effective teacher appraisal is an important element in the drive to improve teaching and learning and raise education outcomes.

This chapter will explore commonalities and differences in teacher appraisal approaches across countries. The analysis throughout this chapter indicates that teacher appraisal is probably the least comparable component of evaluation and assessment frameworks across OECD countries. Evidence from the country-specific *OECD Reviews of Evaluation and Assessment in Education* points to large variations in approaches to teacher appraisal, which range from centralised national systems to informal approaches developed at the discretion of individual schools.

The chapter is organised in eight sections. After this introduction, the second section lays out the analytical approach, followed by a third section on impact, drivers and contextual developments. The following four sections describe key features of teacher appraisal and country practices, structured along the four main topics of the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: governance, procedures, capacity and use of results. The final section provides pointers for policy development.

Analytical approach

Definitions

Teacher appraisal refers to the evaluation of individual teachers to make a judgement about their competencies and performance and/or to provide feedback to support the improvement of their practice. Countries use a range of different approaches for the appraisal, among which the most frequent are: (1) appraisal for the completion of a probationary period; (2) appraisal as part of performance management, which may include registration processes, regular appraisal and appraisal for promotion; and (3) reward schemes. These key terms are defined below.

- **Completion of probation** refers to the appraisal of individual teachers upon completion of a probationary period. The probationary period is a limited period of time upon teachers' entry in the profession during which school leaders and/or other evaluators can closely appraise the competence and progress of newly hired teachers, monitor specific aspects of their performance and identify professional development needs. It is a period of adjustment during which teachers may also benefit from induction and mentoring arrangements. Upon successful completion

of the probationary period, teachers may receive a promotion or apply for fully registered teaching status.

- **Performance management** refers to the formal teacher appraisal processes designed to ensure that individual and organisational goals are met. This includes all types of appraisal related to the management of a teacher's professional and career development. As such, performance management is part of wider processes and systems for measuring, monitoring and enhancing the performance of teachers. It includes processes such as teacher registration, regular appraisal and appraisal for promotion.
  - **Teacher registration**, or certification, processes officially confirm teachers as competent for teaching practice. Advancement to fully registered teaching status typically occurs upon completion of a probationary teaching period and/or following an appraisal against registration/certification criteria. The process typically involves external evaluators or a national agency responsible for teacher registration. After teachers have initially become fully registered, in some countries they have to renew their registration every few years.
  - **Regular appraisal** is typically (but not always) a process internal to the school regulated by general labour-law provisions requiring the employers of teachers to regularly appraise the performance and results of their employees. In some countries, there are prescriptions on how such appraisal should be implemented whereas in other countries school leaders are autonomous in designing their own approaches. The process may be connected to a discussion and plan regarding elements such as the teacher's professional development, responsibilities, working conditions, career progression or salary advancement.
  - **Appraisal for promotion** is a process that is separate from regular teacher appraisal in some countries. It is often voluntary and takes place in relation to decisions on employment status. It should be noted that many countries do not have a specific process for appraisal for promotion but integrate this function into regular teacher appraisal and/or registration.
- **Reward schemes** involve teacher appraisal that is explicitly designed to identify a select number of high-performing teachers to acknowledge and reward their teaching competence and performance through rewards or one-off salary increases.

Conceptual framework

This chapter aims to explore the complex range of features associated with teacher appraisal. Figure 5.1 below provides a conceptual framework summarising the aspects involved and the way they interconnect. The overarching policy objective is to ensure that teacher appraisal contributes to the improvement of student outcomes through improved teaching practices and teacher professional learning. The conceptual framework has four main interrelated aspects.

- **Governance**: This aspect concerns the overall design and organisation of teacher appraisal. This encompasses the purposes of teacher appraisal and the balance between developmental and accountability functions of the appraisal process. It also includes the setting of appraisal requirements as well as the distribution of responsibilities for the design of teacher appraisal.

- **Procedures**: This aspect refers to the features of teacher appraisal and the ways in which these are combined to create a specific appraisal model. This includes requirements for the frequency of appraisal, the development of reference standards, the definition of appraisal aspects and criteria, and the combination of instruments to gather relevant information.
- **Capacity**: This aspect analyses the distribution of responsibilities for the implementation of appraisal as well as the training and support provided to appraise, to be appraised and to use the results of an appraisal. It includes issues such as: the choice of evaluators and the development of skills to perform the appraisal of a teacher; the enhancement of teachers' skills to benefit from their own appraisal; and the development of central expertise in designing appraisal systems.
- **Use of results**: This aspect is concerned with how the teacher appraisal process is followed up and how results are utilised for further decisions or actions. Examples of mechanisms to use appraisal results include performance feedback, professional development plans, career advancement and financial and other rewards.

Impact, drivers and contextual developments

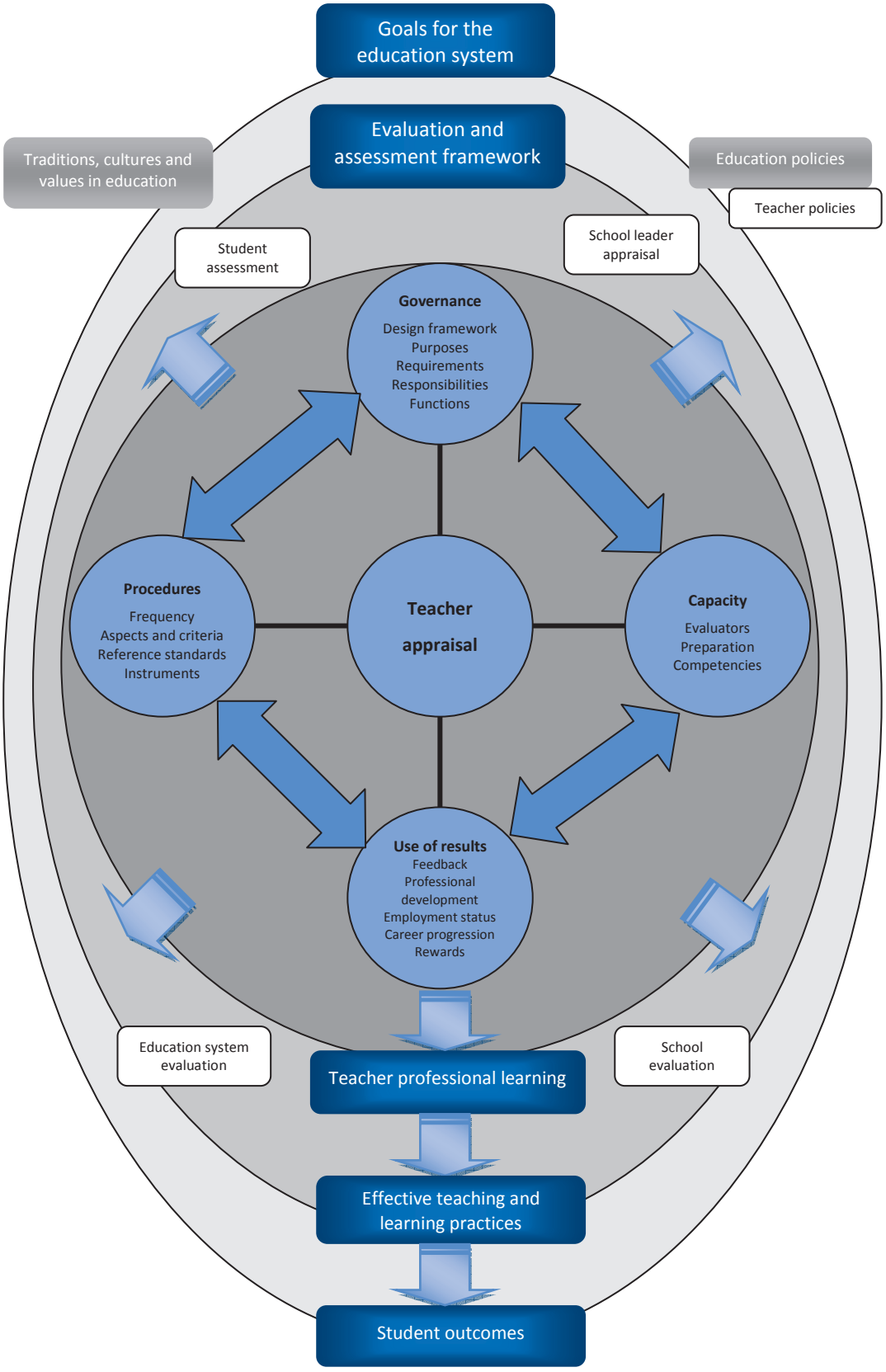
The importance and impact of teacher appraisal

As the most important school-level factor in student achievement, teachers are key to improving education outcomes. Raising the quality and equity of schooling depends to a large extent on making sure that teachers are highly skilled, well resourced, and motivated to perform at their best (OECD, 2005). In turn, the effective monitoring and appraisal of teaching is central to the continuous improvement of schooling (Santiago and Benavides, 2009). It can be a key lever to increase the focus on teaching quality and continuous professional learning for teachers, in line with a widespread recognition of the impact of teaching performance on student learning outcomes.

Effective teacher appraisal also provides a mechanism to recognise and reward high-quality teaching and to manage teacher career advancement. It can facilitate the organisation of schools in ways that are sensitive to individual talent, performance and motivation by allowing teachers to progress in their career and take on new roles and responsibilities based on solid appraisal of their performance (Mead et al., 2012). This can help address concerns about the attractiveness of teaching as a career choice and about the image and status of teachers in a number of OECD countries, including teachers' feeling that their work is undervalued.

Teacher appraisal is also well in line with a general focus on performance appraisal in the public sector and strengthened requirements for evaluation and accountability. In the context of increased parental expectations of schooling and concerns about the allocation of public funding, teacher appraisal provides a tool for schools to be accountable for the quality of education in every classroom. It allows school leaders and education authorities to monitor individual teacher performance and ensure that all classrooms are in the hands of capable and motivated teachers. It can also provide a possibility to move on teachers who are consistently underperforming and not responding to professional development opportunities.

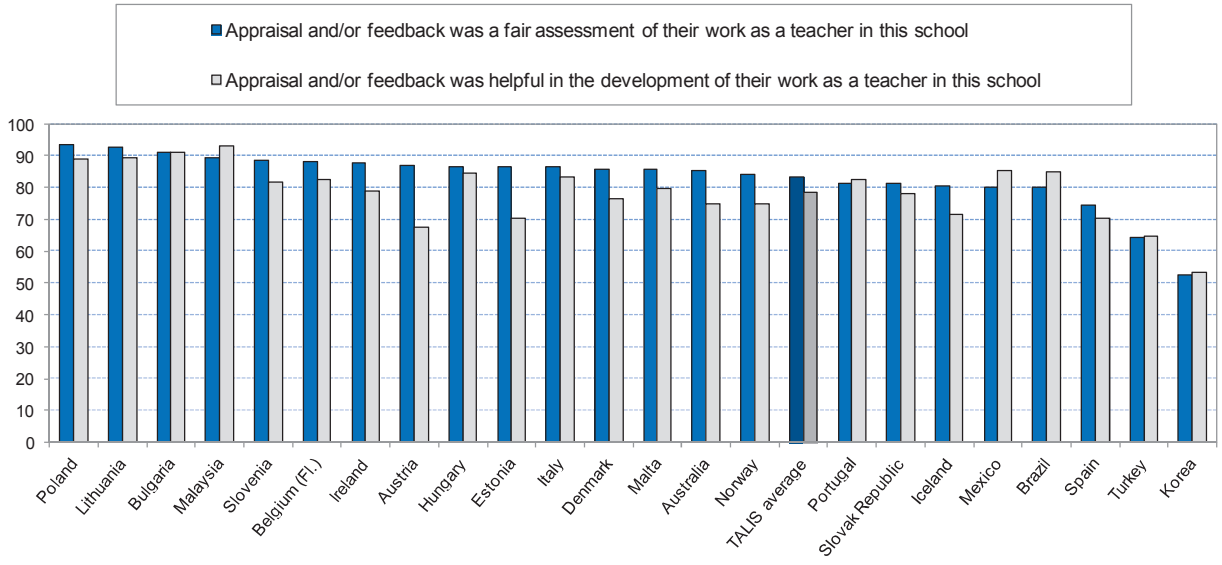
Figure 5.1 Conceptual framework for teacher appraisal



Evidence from the OECD's Teaching and Learning International Survey (TALIS) indicates that teachers at the lower secondary level who received appraisal and feedback generally had positive views of these processes. Overall, a large majority of teachers (83.2% on average across TALIS countries) who had received appraisal and feedback considered it to be a fair assessment of their work and most of them (78.6%) found that it was helpful in developing their work as teachers in the school (Figure 5.2). These are very important findings showing for the first time that representative samples of teachers across countries report largely positive experiences with their own appraisal processes (OECD, 2009).

Figure 5.2 Teacher perceptions of the appraisal and/or feedback they received (2007-08)

Percentage of teachers of lower secondary education who agreed or strongly agreed with the following statements about the appraisal and/or feedback they had received in their school



StatLink http://dx.doi.org/10.1787/888932791305

Note: Only includes those teachers who received appraisal or feedback.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

While measuring the impact of teacher appraisal is conceptually and methodologically challenging, a range of studies conducted in the United States appear to show a positive relationship between certain approaches to teacher appraisal and teacher practices and motivation. Even though research results in this area are more mixed, some studies also point towards a positive relationship between teacher appraisal and improved student learning outcomes (for a brief overview of the research evidence, see Box 5.1). It should be noted, however, that there are huge differences in the design and implementation of teacher appraisal systems across and within countries, which makes it impossible to draw generic conclusions about the impact of teacher appraisal per se. Indeed, it is not the existence of formal appraisal requirements but the design and quality of the processes that seem to matter most for its effect on teaching and learning outcomes.

Box 5.1 How teacher appraisal influences teaching and learning: A brief overview of research evidence

This box aims to summarise research regarding the influence of teacher appraisal on teaching and learning. It is drawn from a literature review on teacher appraisal prepared for the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes by Isoré (2009). It should be noted that this review is based only on evidence from the United States.

The research literature on teaching quality largely establishes that teachers matter for student outcomes, in the sense that teachers are powerful contributors to students' academic achievement (OECD, 2005). However, the literature is more hesitant in demonstrating which teacher aspects are relevant to teacher quality and what is the relative importance of teacher quality *vis-à-vis* other factors that theoretically influence student learning, including family, student and school factors. Not surprisingly, measuring the impact of teachers' appraisal in terms of student learning is even more difficult. Overall, research seems to show that teacher appraisal has an *indirect* impact on student learning, by influencing teacher attitudes and practices, which in turn, can shape student learning outcomes.

There are several strands of research analysing the impact of teacher appraisal systems on teaching and learning. First, there is a body of research which looks at the effects of teacher appraisal on the enhancement of teacher practice and motivation, as perceived by the teachers who are appraised. If teachers report enhanced practices owing to the appraisal process – and assuming that the corresponding practices are relevant to student learning –, then the appraisal system is supposed to be effective at indirectly improving student outcomes. Second, there are a range of quantitative studies attempting to identify a direct link between teacher appraisal and student learning outcomes. Given the many factors influencing student outcomes, such a link is more challenging to establish and the results of this research appear to be more mixed.

The relationship between teacher appraisal and teacher practices and motivation

Several studies have analysed the impact of the voluntary teacher appraisal process proposed by the National Board for Professional Teaching Standards (NBPTS) in the United States because it represents one of the most complex and comprehensive approaches to teacher appraisal and also leads to a formal recognition – the National Board Certification (NBC). Several authors (Bond et al., 2000; Lustick and Sykes, 2006) found that teachers applied in the classroom what they had learned from the appraisal process. Teachers seemed to have also gained new enthusiasm for the profession – regarding how long they plan to stay in teaching – as a result of going through the appraisal process (Vandervoort et al., 2004; Lustick and Sykes, 2006; NBPTS, 2007). Finally, the accomplished teachers who went through the appraisal process were more likely to contribute to school leadership by adopting new roles including mentoring and coaching of other teachers who recognise certified teachers as helpful (Petty, 2002; Freund et al., 2005). Cohen and Rice (2005) concluded that the NBPTS provides a cost-effective opportunity for professional development through the appraisal process by requiring teachers to create portfolios and reflect about their practices, as well as helping them to focus on strong curricula and accurate assessment of student learning. While these studies bring interesting insights regarding formative aspects of teacher appraisal, it is important to keep in mind the potential impact of self-selection, as this is a voluntary appraisal scheme.

The relationship between teacher appraisal and student learning outcomes

Measuring the direct effect of teacher appraisal on student learning outcomes is more challenging. Such research needs to control for the broad set of qualitative variables which are likely to influence student learning. These variables encompass teacher characteristics (e.g. age, gender), teacher education and experience, students' family factors (e.g. socio-economic background, parental support), school factors (e.g. school policies, school incentives, peer and classroom effects) and student factors (e.g. motivation, cognitive abilities, cumulative experience). The complex realities of education prevent researchers from accurately assimilating these factors as traditional inputs into education production functions (Hanushek, 1986). Moreover, because of its qualitative and heterogeneous nature, the output itself – student learning – is not a traditionally measurable "end product", and this makes the decomposition between different factor contributions even more difficult (Hanushek, 1986; Ingvarson et al., 2007). This implies that this type of quantitative study in education requires particular attention to analytical issues or potential misinterpretations of the results.

Box 5.1 How teacher appraisal influences teaching and learning: A brief overview of research evidence (*continued*)

There is a body of research focusing on the variation in the statistical relationship between teachers and student outcomes by comparing teachers who pass a particular appraisal process to teachers who do not. This body of evidence does not assess the effects of teacher appraisal on student outcomes, since it compares two distinct groups of teachers (one subject to appraisal, the other not) instead of comparing the impact of one particular group of teachers on student outcomes before and after the considered appraisal process. Rather, it provides an indication of the capacity of the implemented appraisal process to effectively differentiate proficient teachers from other teachers.

Again, numerous studies examined the appraisal process associated with the National Board for Professional Teaching Standards (NBPTS) in the United States. A number of authors (Cavalluzzo, 2004; Goldhaber and Anthony, 2007; Vandervoort et al., 2004; Smith et al., 2005) found that students of teachers who had obtained the NBC performed better on standardised tests than students of non-certified teachers. Moreover, Goldhaber and Anthony (2007) and Cavalluzzo (2004) also conclude that student scores particularly improved for minority students and special needs students. This suggests that the NBC properly identifies teachers who adopt practices that enhance educational equity in addition to overall quality. However, other authors (McColskey and Stronge, 2005; Sanders et al., 2005; Harris and Sass, 2007) found, by contrast, that students of teachers who had obtained the NBC did not perform significantly better than other students overall, in spite of improvements in some year levels or areas.

The empirical evidence is also mixed for systems of compulsory teacher appraisal. Milanowski (2004) estimated the relationship between teacher appraisal ratings and a measure of value-added student achievement for the US district of Cincinnati, which has implemented a comprehensive standards-based teacher appraisal scheme as a basis for a performance-based pay system. He found significant positive correlations, and concluded that if scores from a rigorous teacher appraisal system are substantially related to student achievement, then this provides validity evidence for the use of the teacher scores as a basis for a financial reward system. Borman and Kimball (2005) studied the teacher appraisal system of the district of Washoe County, with a two-level model. After controlling for student background and teachers' experience, they assessed the relation between teacher quality as measured by the appraisal system and both overall classroom mean achievement and within-classroom effects on social equality. They found that teachers with high appraisal scores were related to better student learning outcomes across year levels and teaching areas (reading and math). But they also found that these teachers did not appear to be reducing gaps in achievement between low- and high-achieving students and students from low-income or minority background.

Finally, a range of studies focus on particular appraisal approaches or instruments. In particular, research has become increasingly interested in teacher appraisal approaches using classroom observations. Two recent studies analysing the relationship between teacher appraisal based on classroom observations and student performance report positive results. The Measures of Effective Teaching (MET) project compares five different instruments for classroom observation and analyses their relationship with a range of student outcomes. The analysis is based on the practice of over 1 300 teachers across different school districts in the United States. The study concludes that all five classroom observation instruments were positively associated with student achievement gains (Kane and Staiger, 2012). With a similar focus, Taylor and Tyler (2011) analyse the effects of the long-running Teacher Evaluation System (TES) in Cincinnati Public schools, which appraises teachers' professional practice through multiple detailed classroom observations and a review of teacher work products. They find that high-quality teacher appraisal based on classroom observation improved the performance of mid-career teachers both during the period of appraisal and in subsequent years. In particular, students assigned to a teacher after he or she had participated in TES scored about 10% of a standard deviation higher in mathematics than similar students taught by the same teacher prior to participation in TES.

Sources: Isoré (2009); Kane and Staiger (2012); Taylor and Tyler (2011).

While a range of studies point to positive effects of teacher appraisal systems (Box 5.1), there is also a body of research analysing the potential negative impact of teacher appraisal systems that are not carefully designed and implemented. Where the accountability function has taken precedence over the developmental function and the appraisal is mostly perceived as punitive, teacher appraisal may create a climate of tensions and fear (O'Day, 2002; Klinger et al., 2008; OECD, 2009). Although accountability is clearly important, excessive focus on this function may jeopardise the key purpose of appraisal, which is to improve teaching quality by developing teacher capacity. In this context, education policy makers, practitioners and researchers in many countries are re-emphasising the primary value of teacher appraisal as a tool for the continuous improvement of teaching practice.

Drivers and contextual developments

Teacher appraisal, like all components of the evaluation and assessment framework, is influenced by wider trends and developments shaping education policies (see Chapter 2). Recent global phenomena such as the growing importance of information and communication technologies (ICT), the increasing impact of migration and social changes on schools, new approaches to school governance and an increasing emphasis on evaluation, assessment and accountability in education have all contributed to changing roles and responsibilities for teachers. This increasing complexity of teachers' roles has raised the importance of providing adequate feedback and support for all teachers to continuously develop their skills. In this context, the (re-)definition of what constitutes good teaching, as well as the appraisal of teaching practices in relation to agreed standards of good teaching, are crucial elements in developing effective teaching for the 21st century. This section briefly explains the key contextual developments impacting on teachers and the teaching profession.

First, the exponential increase in the availability of information over the past decades has fundamentally changed the nature of schooling, with teachers being asked to move away from traditional instructional approaches focused on knowledge transmission. Most OECD curricula now stress the importance for teachers to help their students acquire 21st century skills such as self-directed learning, problem-solving, teamwork and creative thinking, which requires more constructivist approaches to teaching and learning. Teachers themselves are also expected to apply information and communication technologies in their professional practice and use the potential of ICT to improve their teaching approaches.

Second, as school systems are becoming ever more diverse in terms of students' socio-economic, cultural and linguistic backgrounds, teachers are expected to ensure social cohesion and well-being of all students while differentiating their instruction to respond to individual learning needs. In some countries, schools are increasingly offering inclusive education for students with special educational needs, which requires teachers to acquire competencies in this field or to collaborate with specialists. In this context, teachers are also increasingly expected to work beyond the school borders and co-operate with parents and the broader community.

Third, developments in education governance structures towards more decentralisation and school autonomy have led to new managerial tasks in schools, requiring teachers to participate in more widely distributed school leadership and management roles. Also, in line with research on effective organisational learning, many schools are adopting new ways of working that focus on collegial and collaborative teaching, conducted in teams and larger professional learning communities (Stoll and

Louis, 2007). This requires teachers to adapt to collaborative work cultures based on shared goals, continuous professional development, reflective practice, peer observation, feedback and quality improvement.

Finally, the growing prominence of evaluation, assessment and accountability in education policy also has important implications for teachers' work in schools. Many education systems are promoting an increased focus on setting central standards for student learning, measuring student outcomes and using data for improvement. As teaching and school quality are increasingly being judged on the basis of evidence of student learning, teachers need to be able to create valid and reliable assessment instruments to monitor their students' progress and they are also required to collect, interpret and follow up on data from standardised external assessments. On the whole-school level, they are expected to collaborate in school self-evaluation processes to inform school development and communicate assessment and evaluation information to students and parents.

This combination of developments requires teachers of all age groups to develop new competencies which they may not have through formal training. The (re-)definition of professional standards or profiles for the teaching profession can help acknowledge the great complexity of teaching in the 21st century and emphasise the need for continuous learning and development. Teacher appraisal, then, plays a crucial role in identifying individual teacher strengths and learning needs in relation to such expectations and provides a mechanism for teachers to continuously review, reflect on, and improve their practice.

Governance

This section analyses the overall design and organisation of teacher appraisal across countries participating in the OECD Review. This encompasses the purposes of teacher appraisal and the balance between developmental and accountability functions of the appraisal process. It also includes the setting of appraisal requirements as well as the distribution of responsibilities for the design of teacher appraisal.

Purposes

Teacher appraisal can be conducted with different purposes such as completion of a probationary period, teacher registration, performance management, regular feedback for improvement, career advancement, individual inspection or identification of developmental needs. The overarching objectives of these different teacher appraisal processes typically include professional development and/or accountability.

The developmental function

Developmental teacher appraisal, also referred to as appraisal for improvement or formative appraisal, focuses on the provision of feedback useful for the improvement of teaching practices, namely through professional development. It is typically a process internal to the school which may not always be regulated nationally. It has as its main purpose the continuous improvement of teaching practices through adequate professional development. The identification of individual teachers' strengths and weaknesses serves to choose from a wide range of possible professional development activities the ones that meet individual teachers' own needs in relation to the priorities in the school improvement plan. It involves helping teachers learn about, reflect on, and improve their

practice. This typically occurs with account of the school context so that the learning of individual teachers is aligned with school needs.

The accountability function

Teacher appraisal for accountability, also referred to as summative appraisal, focuses on holding teachers accountable for their performance associating it to a range of consequences for their career. It aims to provide summary information about past practices and performance of a teacher relative to what is considered as standards of "good" teaching. It seeks to set incentives for teachers to perform at their best. It may involve external evaluators and typically entails consequences for teachers such as performance-based career advancement and/or salaries, bonus pay, or the possibility of sanctions for underperformance. Teacher appraisal for accountability is summative in nature and usually involves appraising performance at nodal points in a teacher's career. It also works as a means to provide recognition to teachers.

Policy frameworks for teacher appraisal

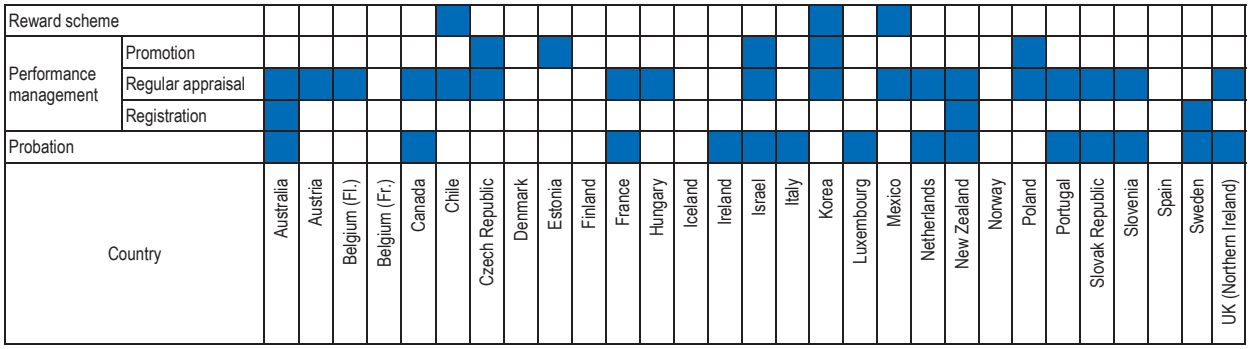
Teacher appraisal is receiving considerable policy attention across countries participating in the OECD Review. Of 29 countries for which information is available, 23 reported having policy frameworks (national or state laws or regulations) in place to regulate one or more types of teacher appraisal. The six education systems that did not have such policy frameworks in place in 2012 were: the French Community of Belgium, Denmark, Finland[1], Iceland, Norway and Spain.

However, the absence of policy frameworks for teacher appraisal does not mean that teachers do not receive any professional feedback in these countries. In Denmark, for example, while teacher appraisal is not conducted systematically at a national level, teachers typically receive appraisal or feedback from their school leaders once a year. In Norway, teacher appraisal approaches are not regulated nationally, but are typically designed at the local and/or school level. In Iceland, teacher appraisal is conducted at the discretion of individual schools and school boards.

The types of appraisal regulated by existing policy frameworks vary across countries. The teacher appraisal approaches most frequently in place across the 29 countries for which information was available are depicted in Figure 5.3. The most common approaches are appraisal for the completion of a probationary period (15 countries) and regular school-based appraisal (18 countries). In addition, as part of performance management, a few countries had specific processes for teacher registration (3 countries) and appraisal for promotion (5 countries) in place. Only 3 countries had policy frameworks for teacher reward schemes.

In addition, a few countries report having teacher appraisal approaches in place that do not readily fit the categories defined above (hence, they are not included in Figure 5.3 above). In Austria, for example, there is an appraisal type for teachers on fixed-term contracts in public and government-dependent private schools only. These teachers, who do not have a permanent position, are bound to re-apply for contract renewal every year during their first five years on the job. In this context, an appraisal occurs annually and is typically conducted by the school principal. A second evaluator, typically an inspector from the state or regional school administration, may be involved in case of a dispute or appeal in the appraisal process. Austria also reported that there is a policy framework for appraisal at the completion of probation, which applies only to teachers on fixed-term contracts at ISCED level 2 (academic programmes only) and 3 (all programmes).

Figure 5.3 Existence of policy frameworks for teacher appraisal across countries (2011-12)



Notes: Appraisal for promotion refers to appraisal schemes that are designed specifically for the purpose of making decisions about promotion. However, regular appraisal may also influence decisions about promotion, in addition to other purposes.

Source: Derived from information supplied by countries participating in the project.

In Luxembourg, in addition to teacher appraisal at the completion of probation, there is also an appraisal process for teachers at ISCED level 1 when a teacher requests to transfer to another school. In the Netherlands, there are two types of appraisal on application and entry into the teaching profession that are not related to completion of probation. These appraisals basically involve administrative checks of whether teachers have the required diploma, no criminal record and no record of improper behaviour. In addition, the Netherlands report that the existing teaching standards may also be used for registration and reward schemes, but currently there is no obligation for schools to implement such processes.

In Poland, in addition to regular appraisal for performance management, there can also be a type of appraisal that may be undertaken at the request of the concerned teacher, the local government, the educational superintendents, the school board or the parents' council. Such appraisal is conducted by the school principal according to procedures prescribed by law and it involves the evaluation of all aspects of the teachers' performance. A negative appraisal leads to the termination of the employment contract; otherwise the appraisal may influence professional development and/or salary decisions.

Some countries have more than one process for teacher appraisal at the end of probation and/or for regular appraisal of teachers. In Canada, while processes vary across jurisdictions, there are typically two processes related to probation and two processes for regular appraisal. Regarding probation, in some provinces/territories, one appraisal takes place during probation in relation to decisions on employment status, and another takes place at the end of the probationary period and for new hires during the first year. For regular appraisal, there is a process for performance management for experienced teachers which takes place once every five years (or in case of performance concerns) in addition to regular appraisal for professional development.

In Mexico, there are several types of teacher appraisal for performance management. In addition to the recently introduced Universal Appraisal System, which is for diagnosis and professional development, there is another regular performance management system for teachers in upper secondary public schools only (*Sistema en Línea para la Evaluación Docente* [SLED]). In addition, there is the opportunity for teachers to take different types of teacher tests on a voluntary basis. The National Exam of Teaching Knowledge and Skills can be taken for entry into the profession and in relation to decisions on employment status (such as applying for a permanent contract). In addition the National

Examinations for the Continuous Training of In-Service Teachers (ENAMS) aim to provide a diagnosis of teachers' professional competencies and to assist teachers in identifying their professional development needs. However, it should be noted that it has stakes for those teachers who are part of the rewards scheme (National Teacher Career Programme [PNCM]), since it is an input for this programme. In this context, the ENAMS serve as a mechanism to "certify" the continuous training that teachers have undertaken.

The importance of the governance context

The implementation of teacher appraisal across schools depends very much on the governance context in place in each country, in particular levels of decentralisation and school autonomy. In several countries that do not have central frameworks for teacher appraisal in place, local authorities and schools have long been in charge of developing local teacher appraisal policies, without much involvement of the central level. But in the context of increasing concerns about variations in the quality and equity of learning across and within schools, several countries have passed reforms to enhance teaching quality and some of these reforms have included a stronger focus on teacher appraisal. In Sweden, for example, where there has been little tradition of teacher appraisal, a new system for teacher registration was introduced in 2011.

National frameworks and requirements for teacher appraisal may create challenges in education systems with a strong tradition of school autonomy in educational decision making. Education authorities need to consider different options to establish the right balance between central guidance for consistently effective teacher appraisal and local flexibility to adapt processes to school needs. For example, if a school or local authority has already made substantial investments in building capacity for a particular teacher appraisal framework and method, requiring it to adopt a national rubric for appraisal may be counterproductive (Mead et al., 2012). On the other hand, in the absence of central guidance, there is a risk that schools develop their appraisal systems in isolation and that local standards and criteria may be too limited in expectations in relation to national education goals. While leaders at the local and school level typically have a better understanding of the schools' specific needs, the involvement of a more central level can allow for system learning and sharing of expertise and good practice. Such tensions between central prescriptions and local flexibility are inherent to the structure of many education systems.

The challenge in decentralised systems is to hold schools and local authorities accountable for the implementation of effective quality assurance policies without stifling the creativity and innovation of local actors in building such systems and making them work in line with local needs. In countries where teacher appraisal is in the purview of local or school-level actors, the central authorities may still play an important role in assisting local actors with the implementation of processes. The role of the central level in decentralised systems may include the development of central teaching standards, requirements regarding the overall mix of instruments and methods to be used (without necessarily mandating a specific tool or approach to be applied), and follow-up to ensure that teachers in all schools can benefit from appraisal and feedback.

Building coherent frameworks for teacher appraisal

Recent research highlights the importance of developing systematic approaches to teacher appraisal that support continuous learning for individual teachers throughout their career and for the profession as a whole (Darling-Hammond, 2012). To build coherent

and comprehensive frameworks for teacher appraisal, many countries combine two or more different processes for appraisal to respond better to different stages of the teaching career and/or to achieve different functions (see Figure 5.3).

For beginning teachers, just over half of the countries for which information was available had appraisal processes at the end of a probationary period in place in 2012 (Figure 5.3). However, in several countries, it was not mandatory for teachers to undergo a probationary period at the beginning of their careers. Across the countries participating in TALIS, more than 19.2% of new teachers (i.e. teachers with two years or less of teaching experience) in lower secondary schools reported that they had never received appraisal or feedback from any internal or external source, compared to 13% of the experienced teachers (OECD, 2012). In some countries, these figures were considerably higher: 60.3% of new teachers in Italy reported that they had never received appraisal or feedback, as well as 32.1% in Spain and Portugal, 25.6% in Ireland and 24.7% in Iceland (OECD, 2012).

This is problematic because there are indications that at this early stage of the teacher career, it is particularly important to allow teachers to work in a well-supported environment and to receive frequent feedback and mentoring (OECD, 2010, 2012). The completion of a probationary period can in fact be considered as a first major step in the teaching career corresponding to access to the first stage of the career structure. Box 5.2 summarises research and experience from different countries regarding the establishment of probationary periods; this summary is taken from the OECD's work on Improving Schools (OECD, 2010).

Box 5.2 Research and experiences regarding probation

In several higher-performing education systems, beginning teachers are required to undertake a period of probation, by the end of which they must have confirmed their capability or competence in order to secure their license to teach. While such probationary and induction periods normally (for example in England and New Zealand) last for one or two years, in some systems (for example in Boston and Chicago in the United States) they can last for as many as three or four. However long they last, such approaches normally provide a mechanism by which those who are judged to be ineffective may be removed from their posts and from the profession, while those who are able to demonstrate their competence may be given or become eligible for a permanent position.

More generally, a probationary period typically takes place, as is the case in England, alongside or as part of a new teacher induction programme which includes mentoring support, thereby creating opportunities for formative as well as summative assessment and for teacher development (Larsen, with Lock and Lee, 2005). However, some studies suggest that the assessment of beginning teachers should not be carried out by the same people who support their induction and early professional development, since (for example) in such circumstances beginning teachers are less likely to admit to areas of weakness and thus to identify their professional development needs (Hobson, 2009; Abell et al., 1995; Heilbronn et al., 2002; Williams and Prestage, 2002).

Source: Reproduced from OECD (2010).

After completion of a probationary period, a number of countries require teachers to become fully registered as teachers. Such a registration system often involves different registration levels and requires teachers to renew their registration after a specific number of years. In Australia (Box 5.3), New Zealand and Sweden, such registration is mandatory for all teachers.

Registration and registration renewal processes at key stages of the teacher's career typically aim to formalise the principle of advancement on merit and are associated with career opportunities or financial rewards for effective teachers. These processes can provide useful information for accountability, hiring and tenure decisions, promotion opportunities, or, in particular circumstances, responses to underperformance. The results are typically also expected to inform future professional development.

Box 5.3 Appraisal systems for registration and certification in Australia and the United States

Teacher registration in Australia

Registration is a requirement for teachers to teach in Australian schools, regardless of school sector. All states and territories have existing statutory teacher registration authorities responsible for registering teachers as competent for practice. The levels of teaching accreditation vary according to the jurisdiction. In most jurisdictions, teachers reach the first level of accreditation from the relevant authority upon graduation from an approved initial teacher education programme. Currently, each teacher registration authority has its own distinct set of standards for registration/accreditation, however, from 2013 jurisdictions will be progressively introducing the Australian Professional Standards for Teachers[2] (the Standards) which will provide a national measure for teachers' professional practice and knowledge. Advancement to full registration (or professional competence) is achieved after a period of employed teaching practice and, from 2013, an appraisal against the Standards at Proficient level. In all states and territories, after teachers have initially become registered within their jurisdiction, they must renew their registration. The period of registration varies but is most commonly five years. The main function of the registration process is that of certifying teachers as fit for the profession mainly through the mandatory process of accessing or maintaining "Full/Competence" status – as such, these processes ensure minimum requirements for teaching are met by practising teachers. Registration processes constitute a powerful quality assurance mechanism to ensure that every school in Australia is staffed with teachers with suitable qualifications who meet prescribed standards for teaching practice. At their initial level (provisional/graduate registration), they also provide a policy lever for setting entrance criteria for the teaching profession and, through the accreditation of initial teacher education programmes, strengthen the alignment between initial teacher education and the needs of schools.

The National Board for Professional Teaching Standards in the United States

When applying to the National Board for Professional Teaching Standards (NBPTS) (*www.nbpts.org*), teachers in the United States enter an extensive application process which consists of two major parts: the portfolio of their work including a videotape of a lesson they have taught; and the assessment centre exercises where teachers address a set of questions that relate to the specific content of their field. The assessment is undertaken against detailed teaching standards established by NBPTS. These are based on NBPTS' five core propositions: *(i)* teachers are committed to students and their learning; *(ii)* teachers know the subjects they teach and how to teach those subjects to students; *(iii)* teachers are responsible for managing and monitoring student learning; *(iv)* teachers think systematically about their practice and learn from experience; and *(v)* teachers are members of learning communities. The standards are developed and reviewed by teachers and other experts. The NBPTS is designed to consider a wide range of teacher competencies, using videos submitted by the teachers to appraise classroom practice and along with portfolio entries focused on teaching practice and constructed response assessments of content knowledge. Submitted materials are reviewed by trained teachers who are experts in the teachers' content areas. In the United States, the NBPTS has been the chief means of certifying that classroom teachers are performing at high levels. It has been considered as a model for other countries who are interested in standards-based certification systems for teachers (Harris and MacKenzie, 2007; Ingvarson and Hattie, 2008). Nearly all states in the United States allow teachers to take the NBPTS examination as a mechanism for increasing their salary, by tying National Board Certification to higher salaries. As of October 2012, the National Board had certified 97 000 teachers nationwide, and more than 6 000 became National Board certified in 2011. The Certification is good for ten years and then the teacher must reapply.

Sources: OECD (2005), National Board for Professional Teaching Standards website, *www.nbpts.org*; Santiago et al. (2011).

In the United States, there is no registration process that all teachers have to go through, but there is a possibility for teachers to participate in a voluntary teacher certification process. If they wish to do so, teachers in the United States can seek national certification through the National Board for Professional Teaching Standards (NBPTS), a privately run but largely government-funded programme. This credential, known as National Board Certification (NBC), is designed to provide recognition to teachers who demonstrate superior knowledge and teaching skills (Box 5.3).

Finally, the most common form of teacher appraisal across countries participating in the OECD Review is regular appraisal as part of performance management. Over two-thirds of the countries for which information is available have such regular appraisal processes in place (see Figure 5.3). These processes are typically organised at the school level and pursue a mix of purposes including professional development and planning of teachers' responsibilities and working conditions. The key aspect of such appraisal is that it feeds into individual and collective professional development. Box 5.4 provides an example from Northern Ireland in the United Kingdom, where regular teacher appraisal is designed to ensure teachers' professional and career development in line with each school's overall improvement goals.

Box 5.4 Using appraisal results for performance review and professional development

Northern Ireland in the United Kingdom has established a Performance Review and Staff Development (PRSD) scheme in 2005, which provides a continuous and systematic process to support all principals, vice principals and teachers (including temporary and part-time) with their professional development and career planning. The components of the review process are the same for all staff regardless of the type of school they are employed in and include three stages:

- **Planning and preparation**: at the beginning of the cycle a reviewee and reviewer(s) meet to discuss and record objectives for the year ahead, reflect on possible outcomes and agree how best to keep progress under review during the year;
- **Monitoring**: this comprises the collection of information relevant to the review and the progress made towards the objectives. It also includes observation of the reviewee in his/her work situation through classroom and/or task observation.
- **Review discussion**: at the end of the review cycle, the reviewer(s) and the reviewee establish the reviewee's performance against agreed objectives and indentify any personal and professional development needs in relation to the agreed objectives. They also agree an action plan and objectives for the incoming year.

The PRSD is closely linked to the school's strategic plan for improvement, the School Development Plan (SDP). The SDP brings together the school's priorities, the main measures it will take to raise standards, the resources dedicated to these and the key outcomes and targets it intends to achieve. It sets out the overall "roadmap" for the three years ahead, with a focus on the school's key priorities and action plans. It is conceived as a living document that all school staff will use as a reference point in evaluating, developing and improving their work. It is the duty of each school's Board of Governors to ensure that training and development needs that are identified through PRSD are reflected in the SDP and that corresponding opportunities for professional development are made available to all teaching staff.

Source: Department of Education, Northern Ireland (forthcoming)

While most countries combine developmental and accountability purposes in a single regular appraisal process, a few systems have established several separate appraisal processes to achieve different functions. For example, the Czech Republic, Estonia, Israel, Korea and Poland have established appraisal processes specifically designed to make promotion decisions and Chile, Korea and Mexico have developed explicit reward schemes to compensate high-performing teachers through rewards or one-off salary increases. Korea is a particularly interesting example where three different appraisal processes have been established with clearly distinct functions: *(i)* appraisal for professional development; *(ii)* appraisal for performance management, which feeds into decisions about promotions; and *(iii)* a performance-based incentive system which is connected to special rewards for teachers taking on specific roles and responsibilities (Kim et al., 2010).

Balancing developmental and accountability functions

As described above, only a minority of countries have separate teacher appraisal approaches for developmental and accountability functions. Across the countries participating in the OECD Review, most teacher appraisal approaches aim to use results for both formative and summative purposes. Combining both the developmental and accountability functions into a single teacher appraisal process raises a range of challenges. When the appraisal is oriented towards the improvement of practice within schools, teachers are typically open to reveal their weaknesses, in the expectation that conveying that information will lead to more effective decisions on developmental needs and training. However, when teachers are confronted with high-stakes consequences of appraisal on their career and salary, the inclination to reveal weak aspects of performance is reduced, i.e. the developmental function may be jeopardised (Santiago and Benavides, 2009).

Also, the approaches and tools best used to achieve either developmental or accountability purposes are not necessarily the same. Teacher appraisal approaches that aim primarily to serve as a basis for human resource decisions and accountability need to provide defensible and comparable evidence of teacher performance (Daley and Kim, 2010; Papay, 2012). Given the strong summative dimension of such appraisals, there is a need for standardisation and an element external to the school to ensure a reliable and unbiased basis for decision making. On the other hand, teacher appraisal approaches that pursue the key aims of teacher learning and professional development need to be developed in alignment with school contexts and objectives and fully engage the teacher through reflection on practice and rich opportunities for professional dialogue and learning.

The use of self-appraisal, for example, illustrates this point. Self-appraisal is a valuable instrument to promote learning and inform professional development (Santiago et al., forthcoming). However, in order for self-appraisal to have value for teachers, and for the profession, it is essential that teachers be able to conduct their self-appraisal in private, with nothing hinging on the results. Otherwise, teachers would have little incentive to report honestly about any problems they face in their teaching, as this might be used against them in an accountability-oriented process (Santiago et al., forthcoming).

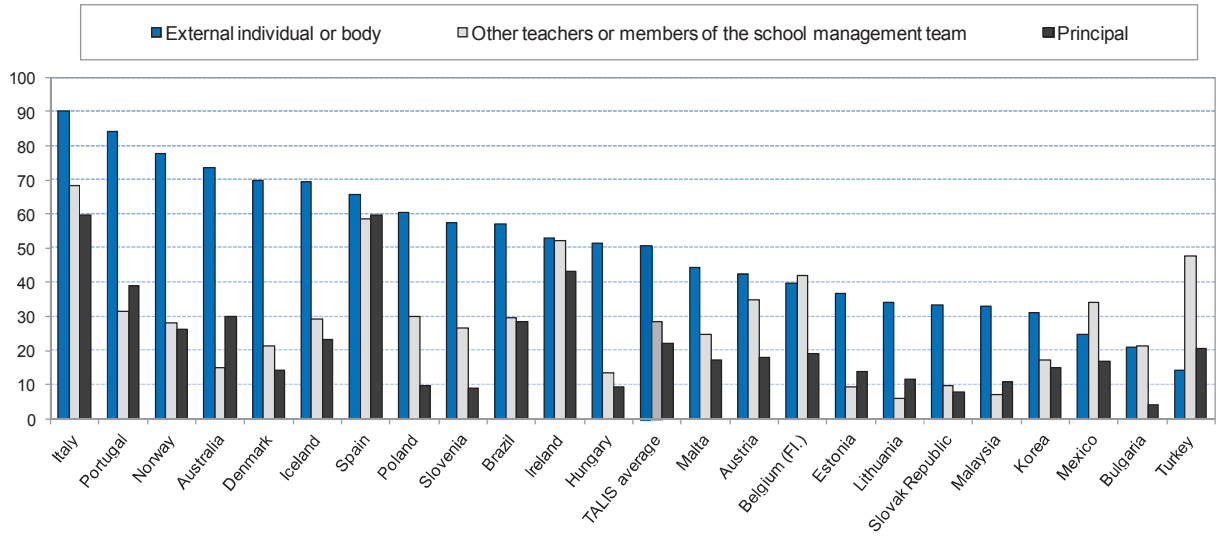
Addressing the challenges of implementation

While the importance of teacher appraisal is widely recognised across OECD countries, many education systems are facing challenges in implementing appraisal for all teachers. As evidenced by TALIS, not all teachers are systematically receiving appraisal and feedback from their employers. On average across TALIS countries, just over half of the surveyed teachers reported that they had never received any appraisal or feedback

from an external source (such as an inspector), which limits these teachers' possibilities to receive a validation of their work by an external entity. Internal appraisal was more frequent across countries, but nonetheless 22.0% of teachers indicated that they had never received any appraisal or feedback from their principal, and 28.6% had never received feedback from other teachers or members of the school management team. Overall, on average across TALIS countries, 13.4% of teachers had never received any feedback or appraisal of their work in the school from any source. These teachers are missing out on the opportunity to receive professional advice from their colleagues and supervisors and may be less well placed than others to engage in focused professional learning and continuously improve their practice (Figure 5.4; OECD, 2009).

Figure 5.4 Teachers missing out on appraisal and feedback opportunities (2007-08)

Percentage of teachers of lower secondary education who reported that they had never received any appraisal and/or feedback from the following sources



StatLink http://dx.doi.org/10.1787/888932791324

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

As the focus on formal teacher appraisal is relatively new in most countries, teacher appraisal appears to be one of the more controversial elements of evaluation and assessment frameworks. Several countries have faced serious difficulties in implementing national teacher appraisal systems. Implementation difficulties may arise as a result of a wide range of factors. First, there might be a defensive culture of evaluation and little tradition of evaluation, feedback and sharing of teaching practices. This might lead to some resistance to application of teacher appraisal by particular groups in the school system. Second, there might be important organisational and capacity challenges to implementation, including: limited professional expertise of those with responsibility to appraise; insufficient preparation of schools to implement appraisal procedures; limited understanding by teachers of the purposes and use of appraisal; a sense of unfairness by those teachers being appraised; an excessive workload inherent to the appraisal process; and a reluctance of many teachers to accept the legitimacy of the evaluators. Third, there might be issues of

lack of resources for specific aspects of the teacher appraisal procedures, particularly the time needed for developmental work, observational evaluation and feedback.

Given these experiences, there is a particular need in the field of teacher appraisal to design policy frameworks carefully and consider the potential unintended consequences. If they are well designed, systems of teacher appraisal and feedback can be powerful levers to increase teacher effectiveness and achieve better student learning outcomes. However, if they are not linked to better classroom teaching and teacher development, teacher appraisal processes may become mere administrative exercises with little impact on education outcomes (OECD, 2009; Jensen and Reichl, 2011). Also, if teacher appraisal approaches focus excessively on the individual teacher, this may increase competition and hamper the teamwork, peer learning and sharing of good practice that are so important to school development and school improvement.

Responsibilities for teacher appraisal

Approaches to appraisal are embedded within the overall structures of governance in education and often depend on the education authorities responsible for teachers as their employer. In the countries that have central (national or state) policy frameworks in place for teacher appraisal, the requirements for appraisal are typically set by the central education authorities. However, as outlined below, a range of other groups may also be involved in setting the rules and procedures for appraisal processes. For a comparative overview of the actors involved in teacher appraisal in each country, see Table 5.1 below.

The role of education authorities

Central education authorities (i.e. the public educational administrations at the national, state or provincial level) play a major role in the conception and application of teacher appraisal in most countries, since they typically set student learning objectives, agree on standards for the teaching profession and establish the norms that regulate teacher appraisal. In some countries, central education authorities play a direct role in the implementation and monitoring of teacher appraisal procedures. This might include the design of specific appraisal tools and instruments, the determination of appraisal criteria, the distribution of appraisal duties, and the follow-up on appraisal results. In other countries, education authorities establish general principles and guidelines only and give schools considerable freedom to adapt the teacher appraisal model to their particular circumstances.

Across the countries participating in the OECD Review, the central/state education authorities or governments are in charge of determining the procedures for teacher appraisal in almost all countries for which information is available, but they often share this responsibility with other actors (Table 5.1). For example, the central/state authorities share responsibility for setting appraisal procedures with the school or school board level in the Czech Republic, Hungary[3], the Slovak Republic (for regular appraisal), Slovenia (for regular appraisal) and Portugal, and with the individual evaluators in France (for regular appraisal). They share this responsibility with teacher professional organisations in Australia (for registration), Ireland (for probation) and New Zealand (for probation and registration), with the teacher union in Mexico (for reward schemes), and with a central agency in Sweden. In Northern Ireland in the United Kingdom, procedures are set by the Teachers' Negotiating Committee, which comprises the Department of Education, employing authorities and teacher unions. In the Netherlands, while central regulations act as a framework, the school organising bodies (competent authorities), represented by

two National Councils of School Boards, act as central employers and are in charge of setting the terms of employment, which schools can further adapt to their context.

The local education authorities are not typically involved in determining teacher appraisal systems in countries that have formal frameworks for teacher appraisal. In some countries with highly decentralised education, such as Denmark, Norway and Sweden, the local authorities may set their own local policies for teacher appraisal, but often they delegate personnel matters, including the appraisal of teachers, to individual school leaders (Nusche et al., 2011a, 2011b; Shewbridge et al., 2011)

The role of inspectorates

The role of inspectorates in determining the procedures for the appraisal of individual teachers is typically limited. Most of the time, the inspectorate does not take responsibility for individual teacher appraisals but, instead, has an important role in stimulating both the quality of school leadership and the quality of teaching. This is typically done through feeding back the results of external school evaluation, mostly consisting of feedback on leadership and management, feedback on the quality of the teaching and learning processes, and feedback on school climate (see Chapter 6). As can be seen from Table 5.1, across the countries participating in the OECD Review, the Inspectorates are not involved in determining the procedures for teacher appraisal. However, the inspectors may be involved in the process as an evaluator in some countries (see section on "Capacity").

The role of teacher professional organisations

In some countries, teacher professional organisations take a lead role in determining procedures for teacher appraisal and participating in teacher appraisal processes. Such involvement of teacher professional organisations or teaching councils is essential to ensure that appraisal processes are relevant for the teaching profession and to create ownership of such processes by the profession. Across the countries participating in the OECD Review, Ireland reports that the procedures for the completion of probation are set by the central education authority in consultation with the national Teaching Council, and in Australia and New Zealand, the teacher professional organisations are involved in teacher registration processes. In Mexico, the teacher union is involved in determining the rewards scheme for teachers. In Northern Ireland in the United Kingdom, teacher unions are represented in the Teachers' Negotiating Committee, which sets procedures for the Performance Review and Staff Development Scheme.

The role of schools and school leadership

The role of school leaders in teacher appraisal differs across countries. In some countries, it may consist of the simple implementation of centrally set regulations. In this case, strong pedagogical leadership is necessary to use the teacher appraisal process developmentally and avoid the image of a bureaucratic device. In other countries, school leaders take full responsibility for setting teacher appraisal processes. For example, in Finland, while there are no national policy frameworks for teacher appraisals, school principals are seen as the pedagogical leaders of the school, responsible for the teachers in their school and for the implementation of measures needed to enhance teaching quality. As a result, most Finnish schools have a system that includes annual discussions aimed at appraising the teacher's fulfilment of individual objectives set up during the previous year and determining developmental needs for the following year.

Among the countries that have formal frameworks for teacher appraisal (Figure 5.3), the school level typically plays a limited role in determining the procedures for teacher appraisal. For the completion of probation, the school organising bodies are involved in determining procedures for appraisal in the Netherlands. For regular appraisal for performance management, schools play a part in determining the procedures in the Czech Republic, Hungary, Slovenia and Portugal. In Hungary schools are legally obliged to set rules for regular appraisal. In Poland, the school principal and school board hold full responsibility for determining performance management procedures. In the Netherlands, the school organising bodies are involved in setting procedures within the central framework. In the Czech Republic, the school principal is also involved in setting procedures for appraisal for promotion.

Table 5.1 Responsibilities for determining procedures for teacher appraisal (2011-12)

Country	Probation	Performance management			Rewards
		Regular appraisal	Registration	Promotion	
Australia	State education authorities or governments; school board or committee (in Independent schools)	State education authorities or governments	Teacher professional organisation; state education authorities or governments	a	a
Austria	a	Central education authority	a	a	a
Belgium (Fl.)	a	Central government	a	a	a
Belgium (Fr.)	a	a	a	a	a
Canada	Provincial/territorial education authorities or governments	Provincial/territorial education authorities or governments	a	a	a
Chile	a	Central education authority or government	a	a	Central education authority
Czech Republic	a	Central education authority or government; school principal	a	Central education authority or government; school principal	a
Denmark	a	a	a		a
Estonia	a	a	a	Central education authority	a
Finland	a	a	a	a	a
France	Central education authority (by ministerial order)	Central education authority; central government; individual evaluators	a	a	a
Hungary	a	Central education authority or government; schools	a	a	a
Iceland	a	a	a	a	a
Ireland	Central education authority in consultation with the teacher professional organisation (Teaching Council)	a	a	a	a
Israel	Central education authority	Central education authority	a	Central education authority	a
Italy	Central education authority	a	a	a	a
Korea	a	Central education authority	a	Central education authority	Central education authority
Luxembourg	Central education authority	a	a	a	a
Mexico	a	Central education authority	a	a	Central education authority or government; teacher union

Table 5.1 Responsibilities for determining procedures for teacher appraisal (2011-12) (*continued*)

Country	Probation	Performance management			Rewards
		Regular appraisal	Registration	Promotion	
Netherlands	Central employer (National Council of School Boards); School organising bodies (competent authorities)	Central employer (National Council of School Boards); school organising bodies (competent authorities)	a	a	a
New Zealand	Teacher professional organisation (The New Zealand Teachers Council)	Central education authority or government	Central education authority or government (The New Zealand Teachers Council is responsible for registering teachers as competent for practice)	a	a
Norway	a	a	a	a	a
Poland	a	School principal; school board or committee	a	Central education authority or government (general framework set by law)	a
Portugal	Central education authority; schools	Central education authority; schools	a	a	a
Slovak Republic	Central education authority (through Act on Pedagogical Employees; Decree of the Ministry of Education)	Central education authority (through the Act on Pedagogical Employees); school principals	a	a	a
Slovenia	Central education authority	Central education authority; school principal	a	a	a
Spain	a	a	a	a	a
Sweden	Central education authority and government	a	Government and central agency (National Agency for Education)	a	a
United Kingdom (Northern Ireland)	Teachers' Negotiating Committee (employing authorities, Department of Education and teacher unions)	Teachers' Negotiating Committee (employing authorities, Department of Education and teacher unions)	a	a	a

Notes: a = not applicable; m = information missing.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Procedures

This section analyses the features of teacher appraisal and the ways in which these are combined to create a specific appraisal model. This includes requirements for the frequency of appraisal, the development of reference standards, the definition of appraisal aspects and criteria, and the combination of instruments to gather relevant information.

Obligation and frequency of appraisal

Teacher appraisal may be mandatory or voluntary, which typically depends on the purpose for the respective appraisal process. The frequency of appraisal also varies across countries. It can be carried out periodically at regular intervals (e.g. every two years), at key stages of the career (e.g. for promotion within the career), or on specific occasions such as at the end of the probationary period or for contract renewal.

Annex 5.A (Tables 5.A.1, 5.A.2 and 5.A.3) provides an overview of country practices. Teacher appraisal for the purpose of completing a probationary period is typically mandatory for all teachers undergoing probation. In most countries, the appraisal occurs at the end of the probationary period. However, in Ireland, it occurs twice during the probationary period in primary schools, and at least three months before the end of the probationary contract in secondary schools. In Australia, the appraisal is generally ongoing throughout the one-year probationary period and varies in frequency. In the Slovak Republic, the periodicity of teacher appraisal for the completion of probation is set at the school level. Where countries have specified the duration of the probationary period, the length was one year in the majority of countries.

Regular teacher appraisal for performance management is mandatory in all countries that have formal frameworks for such processes in place. It occurs periodically on a set cycle in most countries. In Austria and the Czech Republic, however, it is non-periodic and occurs at the discretion of the school principal or the school board. In the Czech Republic, it may also occur in relation to a decision on employment status. In some provinces/territories in Canada, such appraisal may also occur at the discretion of the school leader in case of performance concerns. Regular appraisal for performance management purposes occurs every year in Korea, Mexico, New Zealand, Poland, the Slovak Republic, Slovenia and Northern Ireland in the United Kingdom. In Australia, while there are some variations between states and territories, the introduction of the Australian Teacher Performance and Development Framework will see an annual appraisal process for all teachers. Appraisal for performance management occurs every second year in Portugal, about every second year in Hungary (according to individual school regulations), every third year in Israel and every fourth year at the minimum in the Flemish Community of Belgium and Chile[4]. In the Netherlands, national regulations set the frequency of appraisal as every three years in ISCED levels 2 and 3 and every four years in ISCED level 1, but in practice the frequency of appraisal varies across schools. In some provinces/territories in Canada, experienced teachers are formally appraised every five years. In France, the frequency of appraisal for performance management is highly variable when carried out by inspectors (typically every three to four years at ISCED level 1 and every six to seven years at ISCED level 2). The appraisal by school leaders occurs on a more frequent basis (generally once a year at ISCED levels 2 and 3). Teachers may also be appraised at their own initiative or in the case of problems.

For registration processes, procedures vary among the countries where such systematic processes exist. In New Zealand, it occurs at the end of a "registration period". In Australia, states/territories require renewal after a specific time. In Sweden, appraisal for registration occurs once at the end of an introduction period and may occur again in relation to decisions on employment status, at the discretion of school leaders and the National Agency for Education.

Finally, for reward schemes, in Chile and Mexico, teachers may apply for such appraisal voluntarily. In Chile, there is also an appraisal system for groups of teachers (teaching bodies of individual schools), which is mandatory and occurs every year (the National Performance Evaluation System [SNED]). In Korea, the performance-based incentives system is mandatory and occurs every year. Rewards are given to teachers at the end of each calendar year. The period of appraisal for rewards is from 1 January to 31 December and the actual payments are made in the next year.

Reference standards

A fair and reliable teacher appraisal model needs reference standards to appraise teachers relatively to what is considered "good" teaching. Teaching competences and responsibilities should be listed in order to build a comprehensive definition of what teachers should know and be able to do in the exercise of their profession. The main reference documents for teacher appraisal typically are:

- teacher professional profile or teaching standards (general profile of competencies for teachers), including specialised profiles for particular types of teachers (e.g. level of education, subject)
- set of general and professional duties of teachers, including job descriptions
- at the level of the school, school development plan, the internal regulation and the annual activity plan.

Typically, for reference documents to be used consistently in teacher appraisal processes, they include a range of appraisal criteria to determine the level of performance of individual teachers for each of the aspects assessed. This typically implies the development of indicators and/or standardised forms to record teacher performance. An additional aspect is the weighting of the different aspects assessed in order to compute an overall quantitative rating, in case it is part of the teacher appraisal model. The essential basis for good practice in appraisal is the existence of clear criteria which are consistently applied by competent (trained and experienced) evaluators. This involves the development of explicit guidelines about what is expected from professional practice.

As shown in Table 5.2 below, the types of references used for teacher appraisal vary across countries and appraisal approaches. Austria, Ireland and Italy currently do not have any central reference documents that systematically serve to guide appraisal practices, and the Czech Republic indicated that only school-internal references existed for teacher appraisal.

For teacher appraisal at the end of probation, most countries use central (national or state level) standards or norms as the main reference. Some provinces/territories in Canada and Luxembourg use a description of the general and professional duties of teachers as the main reference. In Australia, in addition to national teaching standards (from 2013) and a description of general duties, the teacher code of conduct is also used as a reference. In France, the relevant description of competencies is determined in a Ministerial order (*arrêté*). In Ireland, for primary education, the Inspectorate publishes appraisal criteria for probation and there is also an appraisal template which provides for ratings in relation to main aspects of practice. For lower secondary education, a code of practice is currently being developed by the Irish Teaching Council to serve as a reference for such appraisal. In Portugal, the appraisal is based on the school development plan and evaluation parameters established by each school as well as national evaluation parameters for classroom observation.

For regular appraisal in the context of performance management, the majority of countries use central (national or state level) teaching standards or norms as the reference for appraisal. In some provinces/territories in Canada, Korea, Mexico and Poland, a description of the general and professional duties of teachers serves as the main reference for this purpose. In Australia, from 2013, the Australian Teacher Performance and Development Framework will be used in addition to national teaching standards. In Mexico, the teacher code of conduct is also used as a reference. On the other hand, in the

Czech Republic, Hungary and Slovenia, school level rules, regulations or development plans are used as the main reference for teacher appraisal as part of performance management. In Slovenia, national regulations on promotions are also considered. In France, Portugal and the Slovak Republic, school level development plans, evaluation parameters or projects are also used as a reference, in addition to a set nationally defined norms or standards (in the case of Portugal, for classroom observation only).

Table 5.2 Reference documents used for different types of teacher appraisal (2011-12)

Country	Probation	Performance management			Rewards scheme
		Regular appraisal	Registration	Promotion	
Australia	State teaching standards; a description of the general and professional duties of teachers; code of conduct From 2013: national teaching standards	State teaching standards; a description of the general and professional duties of teachers; code of conduct. From 2013: national teaching standards; Australian Teacher Performance and Development Framework	National teaching standards; state teaching standards; code of conduct	a	a
Austria	a	None	a	a	a
Belgium (Fl.)	a	National teaching standards	a	a	a
Belgium (Fr.)	a	a	a	a	a
Canada (some provinces/ territories)	A description of the general and professional duties of teachers	A description of the general and professional duties of teachers	a	a	a
Chile	a	National teaching standards	a	a	National teaching standards
Czech Republic	a	School internal regulations	a	School internal regulations	a
Denmark	a	a	a	a	a
Estonia	a	a	a	A description of special tasks and roles	a
Finland	a	a	a	a	a
France	National norms and standards (competency framework in form of a ministerial order)	National norms and standards (through decrees and circulars); school development plan or school project	a	a	a
Hungary	a	School internal regulations	a	a	a
Iceland	a	a	a	a	a
Ireland	ISCED 1: Inspectorate's appraisal criteria for probation; appraisal template ISCED 2: none	a	a	a	a
Israel	National teaching standards	National teaching standards	a	National teaching standards	a
Italy	None	a	a	a	a
Korea	a	A description of the general and professional duties of teachers	a	A description of the general and professional duties of teachers	A description of the general and professional duties of teachers
Luxembourg	A description of the general and professional duties of teachers	a	a	a	a
Mexico	a	A description of the general and professional duties of teachers; code of conduct	a	a	A description of the general and professional duties of teachers; school development plan or school project
Netherlands	National teaching standards	National teaching standards	a	a	a

Table 5.2 Reference documents used for different types of teacher appraisal (2011-12) (*continued*)

Country	Probation	Performance management			Rewards scheme
		Regular appraisal	Registration	Promotion	
New Zealand	National registration standards	National registration standards	National registration standards	a	a
Norway	a	a	a	a	a
Poland	a	A description of the general and professional duties of teachers (as stated in laws and regulations)	a	Developmental plan agreed with the school principal	a
Portugal	School development plan; school-based evaluation parameters; national evaluation parameters for classroom observation	School development plan; school-based evaluation parameters; national evaluation parameters for classroom observation	a	a	a
Slovak Republic	Plan for adaptation education; teacher professional standards	Personal development plan; teacher professional standards	a	a	a
Slovenia	None	School development plan; national regulations on promotion	a	a	a
Spain	a	a	a	a	a
Sweden	National teaching standards	a	National teaching standards	a	a
United Kingdom (Northern Ireland)	Performance Review and Staff Development Scheme (PRDS)	Performance Review and Staff Development Scheme (PRDS)	a	a	a

Notes: a = not applicable; m = information missing.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

For registration purposes, Australia (from 2013) and Sweden use national standards as the main reference for registration. New Zealand has specific Registration Standards that are used as a reference for the purpose of completion of probation, registration and performance management. For promotion purposes, only Israel uses national teaching standards as the main reference. Korea uses a description of the general and professional duties of teachers and Estonia uses a description of special tasks and roles. In the Czech Republic and Poland, teacher appraisal for promotion is based entirely on school-internal regulations or developmental plans. Finally, for rewards schemes, Chile uses the national teaching standards and Korea and Mexico use a description of the general and professional duties of teachers as the main reference. In Mexico, the school development plan or school project is also considered as a reference when appraising teachers for rewards.

Developing a shared understanding of high-quality teaching

While most education systems have some requirements for teachers to be appraised, not all of them have national frameworks or standards for the teaching profession. The absence of such a reference framework and performance criteria for teachers weakens the capacity of designated evaluators to effectively appraise teachers and provide constructive feedback for improvement. Some countries rely entirely on initial teacher education to build a common understanding of high-quality teaching. In Finland, for example, all teachers have Master's degrees and a shared understanding of good teaching is grounded primarily on a unified programme for initial teacher education at the universities providing teacher education.

In some countries with decentralised education systems, individual schools or local education systems may develop their own teaching standards and criteria based on local practice. However, to ensure effective teacher appraisal across the whole education system, it is important that all evaluators have a shared understanding of high-quality teaching and the level of performance that can be achieved by the most effective teachers. A fundamental precondition for the preparation of a profile of teacher competencies is that objectives for student learning are well defined. Teachers' work and the knowledge and skills that they need to be effective must reflect the student learning objectives that schools are aiming to achieve.

Involving teachers in the development of professional standards

For the teaching standards to be relevant and owned by the profession, it is essential that the teaching profession takes a lead role in developing and taking responsibility for them. The participation of teachers in designing standards (and procedures) for teacher appraisal is essential to the effectiveness of any appraisal system. Teachers' participation recognises their professionalism, the importance of their skills and experience and the extent of their responsibilities (Hess and West, 2006). Teachers will be more open to being appraised if there are consulted in the process. Hence, there is a need for appraisal system designers to work hand in hand with teacher unions, teacher professional organisations and outstanding teachers from across the system. Experience from Australia and New Zealand provides some examples (Box 5.5).

Box 5.5 Involving the teaching profession in the development of professional standards

In **Australia**, up until 2012 there have been two distinct types of teaching standards. First, each jurisdiction's statutory teaching body developed its own set of teaching standards for the registration of teachers and the accreditation of initial teacher education programmes. Second, a number of education authorities have also developed distinct professional standards for teachers (e.g. South Australia, Victoria, Western Australia) – these generally provided the reference for performance management processes and establish the links to the career structure. A strength in the system has been the extensive involvement of the teaching profession, employers and teacher educators in the development of teaching standards for registration/accreditation. Teaching colleges/institutes as independent statutory bodies provide teachers with professional autonomy and self-regulation and the right to have a say in the further development of their profession. This reinforces the effective use of standards as a lever for the improvement of teaching practices. In this context, a particularly significant development has been the creation of the Australian Institute for Teaching and School Leadership (AITSL), with the ambition to establish a nationally shared understanding of what counts as accomplished teaching and school leadership. A significant achievement since its creation in 2010 has been the development of the Australian Professional Standards for Teachers (formerly known as the National Professional Standards for Teachers), to provide a national measure for teaching practice. The Standards were developed in close consultation with the teaching profession, employers and teacher educators and are currently being implemented across the country.

In **New Zealand**, the New Zealand Teaching Council (NZTC) has developed teaching standards for the registration of teachers – the Registered Teacher Criteria – and for the accreditation of initial teacher education programmes – the Graduating Teacher Standards (which form the basis to provisionally register teachers). The Registered Teacher Criteria describe the criteria for quality teaching that are to be met by all fully registered teachers and guide the learning of provisionally registered teachers. Positive features of the Registered Teacher Criteria are the increased focus on student learning outcomes, including teachers' analysis and use of student assessment information, and the emphasis on the bicultural context of New Zealand. It is significant that the NZTC as the professional body of teachers holds the leading role in defining standards for the profession, with the extensive involvement of the teaching profession, employers and teacher unions. NZTC provides teachers with professional autonomy, a degree of self-regulation and the right to have a say in the further development of their profession.

Sources: Santiago et al. (2011), Nusche et al. (2012).

Aspects appraised

It is essential to define the key domains and aspects that should be appraised in order to establish a shared understanding of what constitutes "good teaching". The determination of these appraisal domains in national standards or profiles is an important step towards defining what are considered the key responsibilities of a teacher and what type of performance is valued.

An important contribution in this area is Danielson's *Framework for Teaching* (1996, 2007), which is articulated to provide at the same time "a 'road map' to guide novice teachers through their initial classroom experiences, a structure to help experienced professionals become more effective, and a means to focus improvement efforts". The *Framework* groups teachers' responsibilities into four major areas: Planning and preparation, instruction, the classroom environment and professional responsibilities. Each of these areas is further divided into components. For example, planning and preparation includes demonstrating knowledge of content and pedagogy as well as demonstrating knowledge of students and designing instructional goals and corresponding methods. Each of these components consists of several elements to appraise. For example, the teacher's knowledge of students encompasses elements such as knowledge of characteristics of age groups, knowledge of students' varied approaches to learning, etc. The key components of the framework are outlined below:

- ***Planning and preparation***: demonstrating knowledge of content and pedagogy, demonstrating knowledge of students, selecting instructional goals, designing coherent instruction, assessing student learning;
- ***The classroom environment***: creating an environment of respect and rapport, establishing a culture for learning, managing classroom procedures, managing student behaviour and organising physical space;
- ***Instruction***: communicating clearly and accurately, using questioning and discussion techniques, engaging students in learning, providing feedback to students, demonstrating flexibility and responsiveness;
- ***Professional responsibilities***: reflecting on teaching, maintaining accurate records, communicating with families, contributing to the school and district, growing and developing professionally, showing professionalism.

This framework has influenced a large number of teacher appraisal systems around the world. Several states and districts in the United States, as well as the province of Quebec in Canada, have adopted customised versions of the *Framework*'s competency model (Isoré, 2009). Another example can be found in the *Professional Standards for Teachers* in England (TDA, 2007). Chile's *Good Teaching Framework* also has the same "architecture" as Danielson's *Framework for Teaching* (Box 5.6).

Tables 5.A.1, 5.A.2 and 5.A.3 (Annex 5.A) show that the four key areas of teachers' responsibility outlined in Danielson's Framework for Teaching are indeed among the aspects most frequently appraised across the countries participating in the OECD Review. Planning and preparation, the classroom environment and instruction are among the elements most frequently mentioned as key areas for teacher appraisal. Other frequently appraised aspects include: the professional development undertaken by teachers, teachers' contributions to school development, and links to external partners or the community. These elements can be seen as part of teachers' broader "professional responsibilities".

Box 5.6 The Good Teaching Framework in Chile

There is a clear definition in Chile of what constitutes good teaching. Teacher performance standards are contained in the Good Teaching Framework (*Marco para la Buena Enseñanza, MBE)*. The MBE is a guide to improving teaching professional practice that can be used both to meet the guiding needs of beginning teachers and to improve the skills of the more experienced. It includes 21 criteria grouped into four dimensions or domains specific to the task of teachers: (1) creating an environment conducive to student learning; (2) teaching for student learning; (3) preparation for teaching: content organisation based on student learning; (4) professional responsibilities. In addition to a shared understanding of standards of practice, each criterion of practice is elaborated by performance levels (outstanding, proficient, basic, poor). These are written in behavioural language, which permits both teachers and administrators to translate the standards into actual events in the classroom, or in instructional planning. At the time of publication of this report, the Chilean Ministry of Education was planning to revise and update the Good Teaching Framework.

The Good Teaching Framework is a derivative of an important movement internationally towards teaching standards, most particularly in the United States but not restricted to that country. In the United States, it was inspired originally by the efforts of several states (for example, Georgia and Florida in the 1980s), and then followed by a national effort orchestrated by the Educational Testing Service (ETS), one of the organisations in the United States prominent in the design of systems to license beginning teachers. ETS has for many decades offered a series of assessments to individual states to use in their licensing procedures; these were updated in the early 1990s under the general name of "The Praxis Series" comprised of Praxis I (to determine a teaching candidate's grasp of the basic skills of literacy and numeracy), Praxis II (administered at the conclusion of a teaching candidate's period of formal education, to ascertain a future teacher's understanding of the content he or she will teach and knowledge of the appropriate pedagogical techniques), and Praxis III (designed to answer the question, in the teacher's first year of teaching, whether in addition to knowing, for example science, the beginning teacher can actually *teach* science). The Praxis III criteria, originally intended for first year teachers, became the foundation of the Good Teaching Framework in Chile. Praxis III also provided the launch-pad of the Danielson *Framework for Teaching* which has been widely adopted by states in the United States as the official definition of teaching for the purposes of teacher evaluation. This framework shares the same "architecture" as Chile's Good Teaching Framework, but was modified to describe the work of all teachers – not only teachers in their first year of practice – and has evolved in the intervening years to reflect advances in knowledge regarding effective teaching.

Sources: Chilean Ministry of Education (forthcoming), Santiago et al. (forthcoming).

The work of a teacher involves considerably more than the pedagogical activities associated with student learning. It is therefore appropriate that teacher appraisal models consider professional responsibilities less directly related to the teaching itself. This recognises the fact that the demands on schools and teachers are becoming more complex and teachers have their areas of responsibility broadened. Some examples are: working and planning in teams; projects between schools; management and shared leadership; providing professional advice to parents; building community partnerships for learning; and participation in professional development (OECD, 2005).

In a number of countries, the aspects appraised include a range of more specific elements. In Chile, teachers' competence in student assessment is appraised as part of regular performance management. In Estonia, appraisal for promotion considers teachers' work efficiency, in addition to fulfilment of qualification requirements. In Korea, student guidance is an important aspect in all types of teacher appraisal. In Mexico, teachers' length of service is also considered for the reward scheme. In the Netherlands, teacher appraisal (at the end of probation and for performance management purposes) is intended to focus on a range of competencies including interpersonal competence, pedagogical competence, subject matter and didactical competence, organisational competence, team co-operation, co-operation with the

external environment, and reflection and development. In New Zealand, a variety of specific elements to be appraised are outlined in the teaching standards in addition to teaching and learning environments, such as professional relationships and values, bicultural partnership, promotion of inclusive learning environments, responsiveness to diverse linguistic and cultural backgrounds and analysis and use of assessment information. In France, at ISCED levels 2 and 3, regular appraisal carried out by school leaders also includes an appraisal of teachers' "way of serving" (*manière de servir*), which includes punctuality, attendance, authority and "radiance" (*rayonnement*), as well as teachers' conformity to national programmes and reforms. In Poland, the intention of performance management is to assess "all aspects of teacher performance". In Australia, in addition to aspects related to the organisation of classroom teaching, professional development and impact on the broader school community, appraisal for registration purposes also covers the teacher's "conduct" and general "competence", against the Australian Professional Standards for Teachers from 2013. In Northern Ireland in the United Kingdom, teachers set three personal/shared objectives covering the areas of professional practice, pupil and curriculum development, and professional development.

In some countries, there is also a focus on teachers' knowledge in different areas. In Chile, for the teacher rewards scheme, there is an appraisal of the teacher's knowledge of discipline and curricular content as well as pedagogical knowledge. In Mexico, subject knowledge is considered as part of regular appraisal through the universal appraisal system. In New Zealand, professional knowledge and practice are aspects assessed at the completion of probation. In Poland, subject area and didactic knowledge is considered as part of appraisal for promotion. In Portugal, both scientific and pedagogical dimensions are considered as part of regular appraisal. In Slovenia, there is an assessment of teachers' knowledge of legislation, the language of instruction and teaching skills. In the Netherlands, as mentioned above, subject matter competence is one of the key areas to be appraised.

Finally, it is important to mention that in some countries there are no central regulations regarding the aspects to be appraised. This is the case for regular appraisal processes in Austria, the Czech Republic and Hungary. In these cases, the specific scope of teacher appraisal is determined by the individual evaluators.

Instruments and information sources

Gathering multiple sources of evidence about teacher practice meets the need for accuracy and fairness of the appraisal process, taking into account the complexity of what a "good" teacher should know and be able to do. A range of instruments and information sources are typically used to appraise teachers. As shown in Tables 5.A.1, 5.A.2 and 5.A.3 (Annex 5.A), most countries draw on a mix of several instruments and information sources in order to appraise different aspects of a teacher's performance. The most frequently used instruments are classroom observation, interview/dialogue with the teacher, teacher self-appraisal and portfolio.

Classroom observation

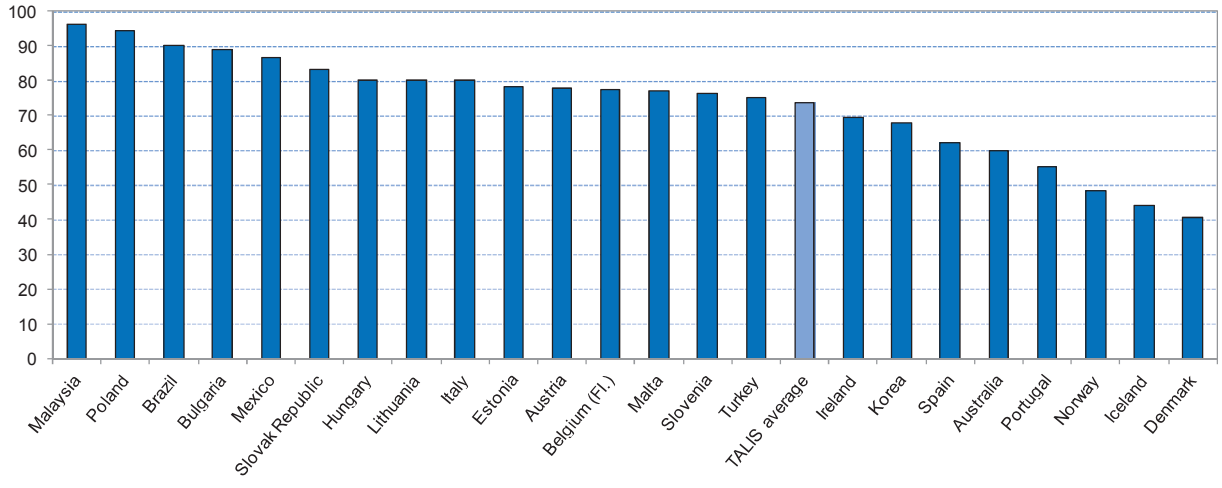
Classroom observations are likely to be among the most relevant sources of information about professional performance, as most key aspects of teaching are displayed while teachers interact with their students in the classroom. Classroom observations appear to be the most common source of evidence used for teacher appraisal, and they are also a key element in school evaluation (see Chapter 6). According to the central appraisal frameworks, almost all countries use classroom observations for

regular performance management and many countries also use it for the completion of probation. Classroom observations are systematically also used for registration in New Zealand and as part of appraisal for promotion in Israel and Korea. In Korea, the performance-based incentive system is also based on observation of performance. While classroom observation is typically undertaken by the school leader or a member of the leadership team, this is not always the case. In Chile, for example, a 45-minutes class is video-recorded and then evaluated in an evaluation centre run by the national institution responsible for teacher appraisal. In Portugal, classroom observations are an optional element of regular appraisal but they are obligatory for the award of Very Good and Excellent marks and for advancement to certain career grades, as well as for teachers who were rated Insufficient on the previous appraisal.

However, while information collected in the OECD Review shows that classroom observations are formally part of most teacher appraisal frameworks, evidence from TALIS also indicates that there are differences across countries in the degree to which classroom observations are perceived by teachers as an important aspect in their appraisal. Especially in countries that do not have a formal framework for teacher appraisal, it is difficult to guarantee that classroom observations consistently take place for every teacher. While on the TALIS average, 73.5% of teachers reported that direct appraisal of classroom teaching was considered with high or moderate importance in their appraisal, this was the case for only 40.7% of teachers in Denmark, 44.1% in Iceland, 58.4% in Norway and 55.3% in Portugal (Figure 5.5).

Figure 5.5 Direct appraisal of classroom teaching as an aspect of teacher appraisal (2007-08)

Percentage of teachers of lower secondary education who reported that direct appraisal of classroom teaching was considered with high or moderate importance in the appraisal and/or feedback they received



StatLink http://dx.doi.org/10.1787/888932791343

Note: Only includes those teachers who received appraisal or feedback.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

In countries where teacher appraisal is more informal and not regulated by central frameworks, school leaders may not routinely be expected to enter classrooms and observe teaching practice with an evaluative focus. Rooting teacher appraisal firmly in

classroom observations can be challenging in countries where there are strong traditions of teacher autonomy and little experience in classroom observations by school leaders or other teachers. In those countries, the main source of professional feedback for teachers is often a dialogue with the school leader regarding issues such as working conditions, responsibilities and salaries without, however, systematically including observation of and feedback on actual teaching practice (Nusche et al., 2011a, 2011b; Shewbridge et al. 2011).

This is problematic because only if teacher appraisal includes classroom observations, it can ensure that individual weaknesses are picked up and robustly addressed with suitable professional development action. Teaching is at the core of a teacher's professional responsibilities, and it can be directly observed. Other proxies of teaching quality such as lessons plans or evidence of communication with parents are of course important pieces of information, but they do not hold the same central position as does the observation of classroom teaching (Santiago et al., forthcoming).

In most countries, classroom observations are conducted by school-based personnel, generally school leaders. Because they are situated in the school, school leaders are (at least theoretically) able to observe classrooms on any day. However, several researchers have criticised the traditional practice of annual announced evaluation visits, which do not provide an authentic picture of day-to-day teaching, are often based on outdated evaluative criteria in the form of checklists and sometimes followed up with simplistic ratings or comments rather than constructive feedback or coaching for improvement (Klinger et al., 2008; Daley and Kim, 2010; Danielson, 2011; Marshall, 2012; Papay, 2012). In several countries participating in the OECD Review there were also concerns about appraisal systems where each teacher was only appraised once every few years.

While high-quality observations appear to be related to increases in student learning outcomes, this relationship is highly dependent on having excellent instruments and well-trained observers (Kane and Staiger, 2012; Kane et al., 2010; Milanowski, 2004, in Santiago et al., forthcoming). Findings from the Measures of Effective Teaching (MET) project in the United States indicate that short observations (15 minutes) are as reliable as longer ones, and that it is important to conduct a number of observations (at least three) to achieve high levels of accuracy (Kane and Staiger, 2012). Also, for classroom observations to be useful for professional learning, each school must have the internal capacity to conduct these accurately. This suggests an extensive investment in training for leadership teams in schools in order for them to conduct effective observations and engage in professional conversations with teachers. Such training should include teachers as well, since it is critical for them to understand how their performance will be assessed. Professional development in this area should focus in particular on helping all those involved gain a deep understanding of what constitutes good teaching practice and how it can be observed (Danielson, 2011).

Objective setting, individual interviews and teacher self-appraisal

Most teacher appraisal models require the individual teacher to set performance objectives for a given period of time in agreement with the school management. The appraisal then assesses the extent to which such objectives were met. The setting of objectives, as well as the appraisal itself, typically involves individual interviews which are an opportunity to trigger critical reflection between evaluators and teachers. Interviews and/or dialogue with the teacher are very frequently used across countries, most often for regular performance management, but also for the completion of probation

and registration processes in a number of countries. In Israel, a dialogue with the teacher is also part of appraisal for promotion. While in most countries, the interview is conducted by a member of the school leadership team, in Chile teachers are interviewed by a peer evaluator (in addition to a written report by the principal).

Another common instrument used in teacher appraisal is teacher self-appraisal. The perspective of the teacher being appraised is essential, because it allows teachers to express their own views about their performance, and reflect on the personal, organisational and institutional factors that had an impact on their teaching. Teacher self-appraisal is used in most countries in regular appraisal for performance management purposes. Teacher self-appraisal is also a formal part of completion of probation in Australia and Israel, of registration processes in New Zealand, and of appraisal for promotion in Estonia and Israel.

Teacher portfolio

Teacher portfolios constitute a powerful tool of teacher appraisal, as they allow to bring together a range of evidence about key aspects of a teacher's competencies and performance. Different elements can compose teacher portfolios, including: lesson plans and teaching materials, samples of student work and commentaries on student assessment examples, teacher's self-reported questionnaires and reflection sheets. It should be noted that portfolios are not only a tool for appraisal per se, but also play a role in supporting a reflective approach to teaching practice. Across the countries participating in the OECD Review, portfolios are quite frequently used for all types of appraisal. In Estonia and France, while the term portfolio is not used, the analysis of different teacher and student documents is part of the appraisal process.

Teacher portfolios are particularly interesting to the extent that they contain artifacts of teacher work which can be differently combined according to the purpose emphasised. On the one hand, several researchers argue that portfolios provide assessment information to hold teachers accountable for meeting educational standards (Klecker, 2000; Campbell et al., 2000; Tucker et al., 2002). On the other hand, Darling-Hammond (2001) argues that teacher development should take precedence in designing portfolios and that "narrative reflection" is the best way to foster such development. Beck et al. (2005) observe that portfolios that focus on teacher development better support professional outcomes. Portfolios are particularly adequate instruments for teacher self-reflection because the proper decision made by the teacher to include particular artifacts (lesson plan, videotape of lesson, sample of student work, narrative comments) instead of others is a judgement that requires determining how the features of one artifact are superior to others (Danielson, 1996, 2007; Darling-Hammond, 2001; Mansvelder-Longayroux et al., 2007). However, combined with other appraisal instruments, documents prepared by the teacher may be used for a summative purpose.

In some contexts, however, the requirement to develop a portfolio can be considered by teachers as a significant additional workload taking their time and attention away from their core work of teaching. Hence, the priority in systems that rely on portfolios should be on instruments that capture authentic teaching practices. For teachers to experience their work on a portfolio as a rewarding experience, they should be designed in a way as to reflect a "natural harvest" of the teacher's work. For example, if teachers have to provide planning documents, these should describe a lesson or unit that the teacher is actually teaching rather than something created specifically for the portfolio. This feature

of "natural harvest" can make the requirement less burdensome to teachers than would be the case if it were perceived as an add-on to their normal responsibilities (Santiago et al., forthcoming).

For example, in the United States, the National Board for Teaching Standards offers certification to teachers who satisfy demanding requirements for a portfolio submission. It is offered in a number of different levels and disciplines (e.g. secondary school mathematics) and requires a total of 13 portfolio entries for a full submission. It is clearly a major undertaking for teachers but it is designed to be directly related to teachers' actual everyday teaching, with portfolio entries expected to be artefacts of their regular practice (Santiago et al., forthcoming).

Student results

Student learning outcomes may reflect some aspects of teaching quality. But student test results are not systematically used as sources of evidence for teacher appraisal in most countries. Across the countries participating in the OECD Review, only the Slovak Republic reports that student outcomes are used for teacher appraisal at the completion of probation, and only Mexico reports that student outcomes (standardised assessment results) are used for regular appraisal in the context of performance management. Mexico also uses student outcomes to appraise teacher performance as part of the rewards scheme. In Chile, students' standardised assessment results are used to appraise groups of teachers (teaching bodies in individual schools) as part of the National Performance Evaluation System (SNED).

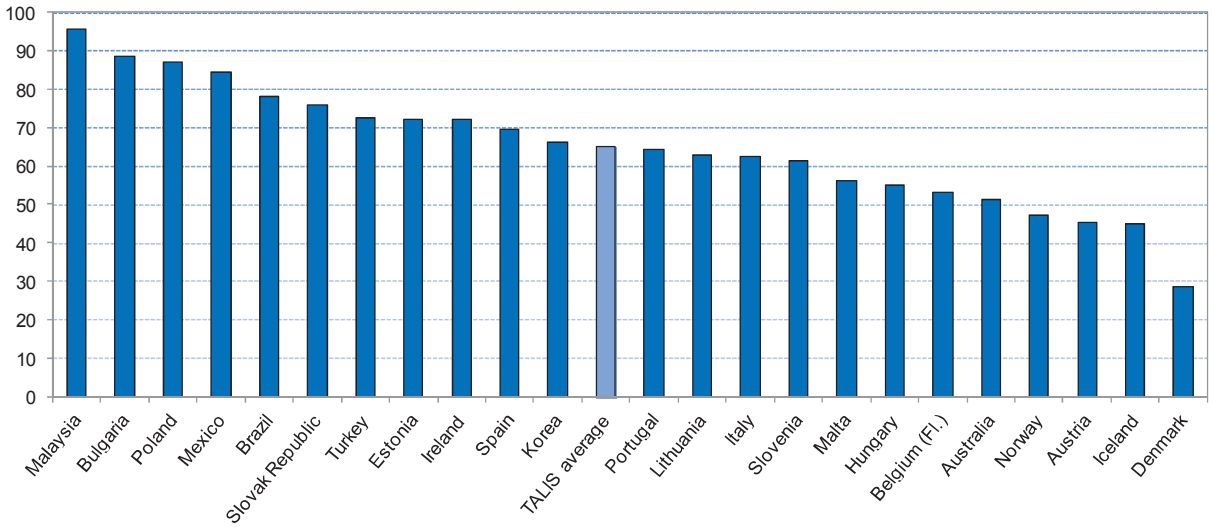
In the United States, there is increasing focus on integrating measures of student growth in teacher appraisal systems. In 2009, the central administration launched the USD 4.35 billion Race to the Top Fund to reward states that created "conditions for education innovation". While the Fund targets several levels of reform, it specifically emphasises the need to improve teacher effectiveness by enhancing teacher appraisal systems. Such appraisal systems are expected to integrate multiple measures of teacher performance including evidence of student learning as a significant element (Hanover Research, 2011).

While student results are not currently used in the formal appraisal frameworks in most countries, in several countries different groups expressed interest in further exploring opportunities to look at student results in order to appraise teachers. Also, in the countries participating in the OECD Teaching and Learning International Survey (TALIS), a considerable majority of teachers (65%) reported that student test scores were in fact considered as an important aspect in the appraisal or feedback that they received (Figure 5.6).

Teacher appraisal systems based on student test results are expected to strengthen incentives for teachers to commit themselves to helping all students meet centrally defined standards and fulfil goals within the national curriculum. Student learning outcomes, including student results in standardised assessments, are an appealing measure to assess teaching performance, since the ultimate goal of teaching is to improve student learning. Braun (2005a) argues that considering student results is a promising approach for two reasons: first, it moves the discussion about teacher quality towards student learning as the primary goal of teaching, and second, it introduces a quantitative and seemingly objective measurement of teacher performance.

Figure 5.6 Student test results as an aspect of teacher appraisal (2007-08)

Percentage of teachers of lower secondary education who reported that student test results were considered with high or moderate importance in the appraisal and/or feedback they received



StatLink http://dx.doi.org/10.1787/888932791362

Note: Only includes those teachers who received appraisal or feedback.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

However, identifying the specific contribution of a given teacher to student outcomes is faced with numerous statistical challenges. A major challenge is that student learning is influenced by many factors. These include the student's own skills, expectations, motivation and behaviour along with the support they receive from their families and the influence of their peer group. In addition to the quality of teachers, other factors influencing student learning include school organisation, resources and climate; and curriculum structure and content. The effect of teachers is also cumulative, i.e. at a given moment in time student learning is influenced not only by the current teachers but also by former teachers. As a result, at a given point in time, raw standardised student scores carry much more than the impact of the current teacher and also reflect, for instance, the impact of the student's family, the student's previous learning or the resources of the school (Isoré, 2009).

In this respect, the development of value-added models represents significant progress as they are designed to control for the individual student's previous results, and therefore have the potential to identify the contribution an individual teacher made to a student's achievement. However, research has pointed to a number of challenges in using value-added scores for the appraisal of individual teachers (based on Santiago et al., forthcoming):

- While value-added models can take a range of variables into account, they cannot control for all the factors which influence student achievement scores other than the teacher's impact (Ingvarson et al., 2007; Goe, 2007).
- Sampling variations can cause imprecisions in test scores, particularly in small schools, where the limited number of students per classroom creates large idiosyncrasies of the particular sample of students being tested (Kane and Staiger, 2002).

- It has also been shown that value-added scores for the same teacher may fluctuate considerably from year to year even though most would expect teacher quality of experienced teachers to be relatively stable (Schochet and Chiang, 2010).
- No single standardised assessment can measure all the curriculum goals for a particular subject, which may put those teachers who focus on teaching all of the curriculum standards at a disadvantage compared to teachers who focus primarily on the standards that are most likely to be tested.
- Value-added models can only be used as a measure of teachers' contribution to student learning growth in subjects where standardised assessments are routinely used to measure students' competencies. While mathematics and reading are tested in most countries, standardised assessments in other subjects are less common. This means that unless additional standardised assessments are created in the non-tested subjects and year levels, a dual appraisal system will be the result – those who are appraised with standardised assessment results and those who are not.
- Value-added models require vast amounts of data to be collected through large scale national-level student testing across levels of education and subjects, and typically require collaboration with an external partner to measure value-added performance, which makes this a costly undertaking (Meyer and Christian, 2008).

Finally, it should also be noted that where high stakes for individual teachers or schools are attached to standardised assessments, distortions in the education process may occur, such as excessive focus on teaching students the specific skills that are assessed, narrowing the curriculum, insufficient focus on key competencies that are not assessed, distributing repeated practice tests, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and sometimes even outright manipulation of results (for detailed reviews of this research, see Morris, 2011, and Rosenkvist, 2010).

Overall, despite the attractiveness of the idea, there are numerous caveats against the use of student standardised assessment results to appraise teachers. In particular, there is wide consensus in the literature around two specific directions: student outcomes should not be used as the *sole* measurement of teacher performance; and student outcomes should not be naively used for career decisions concerning the teacher, including the link to pay, because this incorporates a substantial risk to punish or reward teachers for results beyond their control (Kane and Staiger, 2002; McCaffrey et al., 2003; CAESL, 2004; Braun, 2005b; Ingvarson et al., 2007).

Given that evidence of student learning progress is fundamental, this should not imply that teachers are exempted from providing evidence to demonstrate student progress in their classrooms. For example, teachers could be required to provide specific evidence of student growth and/or portfolios tracking student progress. It is also possible to design a system where teachers and school leaders meet and agree on specific goals for the learning of their students and for ways to assess student progress towards these goals. Such a system would encourage teachers to work with their colleagues and school leaders to identify measurable learning and performance goals for the entire class as well as for groups of students. For example, a teacher with many struggling students may have both a class goal and a goal specifically for the struggling students. In this context, it is important that teachers not be penalised for setting high goals that are not always met

because that might result in teachers setting less-challenging goals for their students. Rather, the students' success and progress, even if they fall short of the goals, should be the basis for measuring teachers' contributions to student learning growth (Santiago et al., forthcoming). Several states in the United States are integrating different measures of student growth in teacher appraisal schemes. Box 5.7 provides some examples.

Box 5.7 United States: Using evidence of student growth to appraise teachers

Rhode Island's teacher appraisal system consists of a sophisticated system to ensure that every teacher is regularly evaluated on both teaching practice and contributions to student learning growth. Regarding the use of student assessment results for teacher appraisal, Rhode Island recognises that not everything that is valued in student learning can be measured with a standardised test. Thus, teachers' contributions to student learning are measured with multiple sources of data on student learning, including portfolios and teacher-made assessments. Even teachers whose students take standardised tests must also examine student learning through other methods. Student learning objectives are used to measure teachers' contribution to student learning growth in all subjects. School leaders approve teachers' objectives and their choice of assessment. Those teaching the same subject and year level within a school meet together and agree on objectives and assessment, which ensures that at least within the school, results for teachers of the same subject and year can be compared. To provide oversight to the schools and ensure that appraisals are carried out in accordance with state policies, each school district within the state must have a District Evaluation Committee that includes teachers, support professionals, administrators and union representatives. The Committee reviews validity and utility of the results from the appraisal process and determines whether the decisions made using evaluation data are fair, accurate and consistent.

Delaware's teacher appraisal system, the Delaware Performance Appraisal System (DPAS), is based on the four domains of Danielson's Framework for Teaching. In 2005, the state started piloting a revised teacher appraisal system (DPAS II), which added student growth as a fifth domain (Component V). DPAS II allows teachers to use student achievement measures based on standardised test scores, internal and external assessments of student achievement and teacher-developed growth goals. Based on the teacher's subject speciality, the state assigns certain assessment measures. To ensure that all measures are comparable and rigorous, the state developed rubrics that help reviewers assess the quality of internal and external assessments. In 2011-12, more than 400 teachers identified assessments that they believed would meet the requirements of DPAS II. These assessments were then reviewed by the Delaware Technical Advisory Group (DTAG) for validity, reliability and rigour. The state also hired a consultant to assist with developing internal assessments across a wide range of subjects. After attending a five-day workshop, cohorts of teachers then created assessments that were reviewed by the DTAG. The state has defined a scheme according to which different internal and external student growth measures will be weighted from 2012-13.

In the state of **Tennessee**, a committee of arts education teachers developed the Tennessee Fine Arts Assessment system which is based on portfolio assessment of students. The assessment is conceived to be used Pending approval from the state board of education, this assessment will be used as an option as a measure of student growth in arts subjects for teacher appraisal in Tennessee from 2012-13. The assessment covers four arts learning domains: perform, create, respond and connect. Teachers use student growth rubrics to determine student progress across these domains. They gather, pre-score and submit a representative sample of student work samples. The samples may include tasks such as visual artwork, written assessments and project-based work. They are then reviewed by a review committee composed of content-specific exemplary teachers to assess student growth towards state standards.

Sources: Bivona (2012); Rhode Island Department of Elementary and Secondary Education, *www.ride.ri.gov*; Santiago et al. (2012).

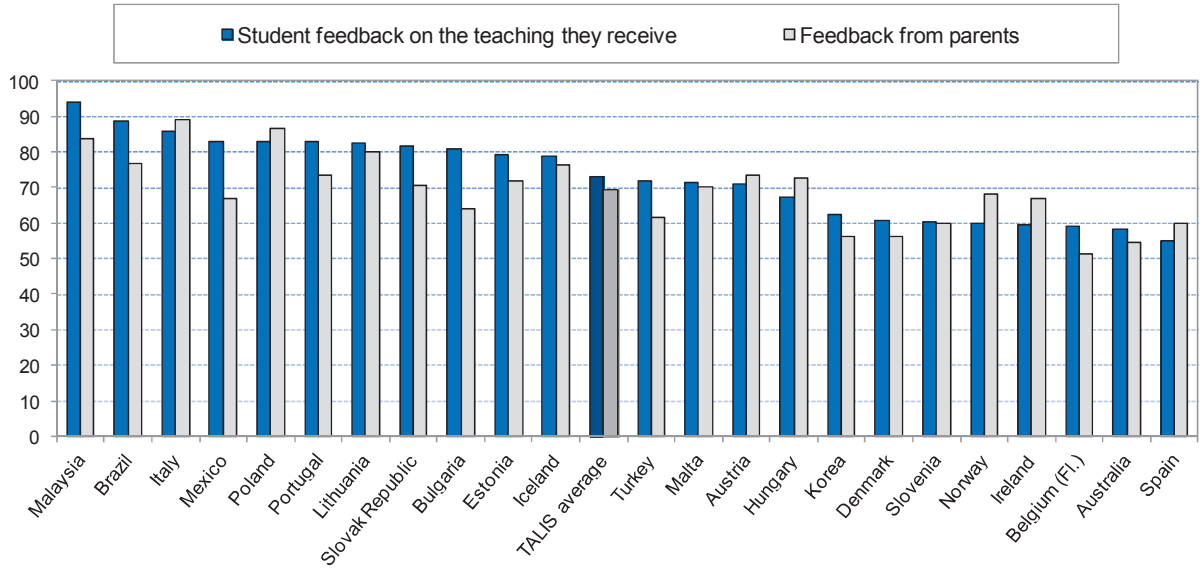
Surveys of students and parents

Surveys can provide information on perceptions of teachers' practice by students, parents and others who may testify on teaching quality through their continuous interaction with the teacher (Peterson, 2000; Peterson et al., 2000, 2003; Jacob and Lefgren, 2005). Surveys collecting the views of students and parents are rarely used systematically in the context of formal frameworks for teacher appraisal. Across the countries participating in the OECD Review, Mexico reports using student surveys for performance management and New Zealand and the Slovak Republic report using student and parent surveys for regular appraisal for performance management in some cases. In some provinces/territories in Canada, parent surveys are used as a component of teacher appraisal. Poland reports that the opinion of the parents' council is considered for the rewards/promotion scheme. In Korea, for regular teacher appraisal for professional development, a multi-dimensional appraisal method is used which involves input from the education community, with students and parents providing information on satisfaction levels through questionnaires.

While stakeholder surveys only rarely form part of countries' formal teacher appraisal frameworks, evidence from TALIS indicates that many teachers across TALIS countries perceive that feedback from students and parents was considered an important aspect of their appraisal (Figure 5.7). Also, in some countries, it is common practice that teachers design their own student surveys to obtain feedback on their teaching practices and their students' learning progress.

Figure 5.7 Student and parent feedback as an aspect of teacher appraisal (2007-08)

Percentage of teachers of lower secondary education who reported that student and parent feedback was considered with high or moderate importance in the appraisal and/or feedback they received



StatLink http://dx.doi.org/10.1787/888932791381

Note: Only includes those teachers who received appraisal or feedback.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Student and parent surveys can provide important formative feedback to teachers (for examples from Norway and Sweden, see Box 5.8). However, it should be noted that while student surveys can yield useful insights, cautions have to be taken regarding the ways in which the results of surveys focussing on individual teachers are used. Quite appropriately, in most contexts, student surveys are not reported to higher levels of the school administration and are generally used only for developmental purposes following the judgement of the concerned teacher. Also, parent surveys are more relevant for whole-school evaluation than for individual teacher performance appraisal. As explained by Isoré (2009), the little current evidence on this subject shows that parents often value teacher characteristics that are not necessarily linked to student achievement, such as "the teacher's ability to promote student satisfaction" (Jacob and Lefgren, 2005), "humane treatment of students" and "effective communication and collaboration with parents" (Peterson et al., 2003).

Box 5.8 Using student feedback as formative feedback for teachers

In **Norway**, in a national-level initiative, the Norwegian Student Organisation and the Union of Education Norway (the largest union for teachers and school leaders in Norway) have been working together in recent years to develop principles and guidelines for teacher appraisal by students. Their aim is to propose a common system that can easily be used and adapted for individual subjects by schools across Norway. They have suggested several features which, in their view, could help ensure that student feedback for teachers will be useful to improve teaching and learning. Not all stakeholder groups agree with the principles that emerged from this co-operation, but the general idea that student views are an important source of feedback for teachers seems widely accepted. The recommendations resulting from this collaboration include that student surveys should: Focus on teaching practice rather than the teacher as an individual; Include the students' own self-assessment and assessment of peers so as to allow for analysis of how student effort and motivation influence the learning environment; Feature questions on teaching approaches that are relevant for student learning such as adapted education and feedback to students as well as questions on the general framework for teaching such as materials and physical conditions; Be carried out anonymously so as to ensure students give honest answers; Be analysed by the teacher and students together with a view to improve the classroom environment and learning outcomes. This should be followed up with a joint report by the teacher and student group on their analysis of results and agreed future changes. This report, together with relevant data, should be submitted to the teachers' closest supervisor.

In **Sweden**, reflecting the student-centred approach to education, teachers often run surveys among their students with the objective of obtaining student feedback on their teaching practices and the learning in their classroom. These surveys are organised on the teachers' own initiative and their results are used exclusively by the concerned teacher often in interaction with the students. Student surveys are kept within the classroom and used only for developmental purposes following the judgement of the concerned teacher. Peterson et al. (2000) argue that students respond reliably about teacher quality if questions are formulated in a simple and relevant way. Teachers interviewed during the review visit for the OECD Review of Evaluation and Assessment in Sweden expressed that students provided useful views into their strategies for teaching and learning and that they found this opportunity for feedback important as a way to consult students on their learning.

Sources: Nusche et al. (2011a, 2011b).

Teacher testing

In some countries, for particular purposes such as the access to a permanent position or entry into the profession, teachers are the subject of testing to assess their general and subject-matter competencies. In some instances, the results of such tests can be used for teacher appraisal. The use of teacher tests or examinations is relatively rare across countries participating in the OECD Review. Chile and Mexico report using teacher tests for their rewards schemes. Mexico also uses teacher tests for regular performance appraisal through the universal appraisal system, and there are several voluntary examinations that teachers can take for entry in the profession and for the diagnosis of their professional competencies. Luxembourg and Slovenia report using a national examination for teachers at the completion of the probationary period.

Other indicators of teacher performance

A number of countries report using appraisal instruments and information sources that do not readily fit the above categories. In France, for the completion of probation, the main instrument used by the specially appointed jury is a report prepared by the teacher's tutor. The inspectors entitled to give their opinion refer to the tutors' reports. At ISCED levels 2 and 3, inspectors also take the opinion of the school leader into account. Candidates for the title of *agrégé* (the most highly qualified category of teachers) are appraised by general inspectors. In Italy, the information used for completion of probation is about formal aspects of a teacher's work, such as presence in school and participation in training. In the Netherlands, an extensive description of a teacher's competencies is available for both completion of probation and regular performance management. In Austria, there are no particular regulations regarding the instruments to be used for performance management; the choice of information sources is up to the evaluators. In Portugal, the instruments used for regular appraisal are outlined by each school in the school development plan. The overall judgement of the school leader informs the application of teachers for registration in Sweden, the regular appraisal for performance management in Hungary and the appraisal of teachers for promotion in Estonia.

Using multiple sources of evidence for teacher appraisal

Drawing on several appraisal instruments and using multiple sources of evidence is advisable for teacher appraisal systems in order to measure different knowledge and skills and obtain a comprehensive picture of teachers' abilities (Goe et al., 2008; Peterson, 1987; Rockoff and Speroni, 2011). This has particular importance when appraisal results are used for high-stakes decisions such as teacher promotion and tenure (Sykes and Winchell, 2010). The Measures of Effective Teaching (MET) project in the United States found that different measures added different and valuable information about teacher performance to the overall appraisal (Kane and Staiger, 2012). In some countries, the use of balanced scorecards has been piloted in a limited number of schools. Such scorecards reflect school and teacher objectives that build on what each school defines as effective teaching and learning (Jensen and Reichl, 2011).

Classroom observation, interviews with the teacher and teacher self-appraisal can provide different perspectives on the teachers' instructional style and teaching approaches. With portfolios, teachers have an opportunity to provide meaningful information that they believe best represents their practice (Attinello et al., 2006; Pecheone et al., 2005; Tucker et al., 2003; van der Schaaf et al., 2008). Student and parent

surveys may capture perspectives missing from other measures (Peterson et al., 2003). Finally, measures of teachers' contributions to student learning growth can add information about teachers' success in helping students master different curriculum areas (Berry et al., 2007; Glazerman et al., 2011; Steele et al., 2010). Box 5.9 provides an example from New York illustrating the use of multiple sources of evidence for teacher appraisal.

While comprehensive teacher appraisal models – e.g. with the multiplication of instruments and evaluators – are more likely to provide a solid basis to appraise teachers, limited resources make trade-offs inevitable. As explained in Isoré (2009), comprehensive teacher appraisal procedures imply greater direct and indirect costs at every stage of the process: reaching agreements on the design of the system requires time for discussions and consultations with all stakeholders; training evaluators is expensive and requires time; conducting appraisal processes creates additional workload for both teachers and evaluators, unless time is made available by reducing workload with other responsibilities; and aligning broader school reforms such as professional development opportunities requires more educational resources.

Box 5.9 Using multiple instruments for teacher appraisal

New York State's teacher evaluation system is typical of most state evaluation systems in the United States in that it provides considerable flexibility to school districts (local education agencies) in selecting instruments for teacher evaluation purposes. However, the state maintains control over the weighting of the multiple measures used in the evaluation process. In addition, the state approves some instruments (observations and surveys) while giving the districts greater discretion in approving measures of teachers' contributions to student learning growth.

The key features of New York's teacher evaluation system are:

- Multiple measures of teacher performance are required for teacher evaluation, including classroom observations and evidence of teachers' contribution to student learning growth (standardised test score growth in tested subjects and year levels along with district-approved measures of student learning growth for all teachers). Student growth measures constitute 40% of teacher evaluation scores while other state-approved measures such as classroom observations, surveys and portfolios constitute the remaining 60%.
- Student learning objectives (SLOs) are used to measure teachers' contribution to student learning growth in all subjects. Teachers receive guidance in setting appropriate learning objectives for their students and districts exercise considerable discretion in approving appropriate assessments and measures to determine student growth.
- Observations must be at least 31% of the 60%, and a minimum of two observations must be conducted each year for each teacher. Anyone conducting classroom observations must be trained and certified to ensure that results of such observations are consistent across classrooms. Districts may select from a variety of state-approved observation protocols (such as Danielson's Framework for Teaching, CLASS, Marzano's Causal Teacher Evaluation System, etc.).
- State-approved parent and/or student surveys may be used as part of the 60%, as well as structured review of lessons plans, portfolios and/or other teacher artifacts.

Sources: New York State Education Department, *www.engageny.org*; Santiago et al. (forthcoming).

Capacity

This section analyses the distribution of responsibilities for the implementation of appraisal, the training and support provided to appraise, to be appraised and to use the results of an appraisal. It includes issues such as: the choice of the evaluators and the development of skills to perform the appraisal of a teacher; the preparation of teachers to be the subject of an appraisal; and the development of competencies to effectively use the results of an appraisal for the improvement of teaching practices.

Internal and external approaches to teacher appraisal

Teacher appraisal might be mostly externally driven or primarily school-based. In the former case, aspects assessed, instruments used as well as appraisal criteria are common across schools and evaluators are predominantly external to the school of the appraised teacher. In the latter case, the school takes responsibility for designing specific appraisal criteria and instruments, following-up appraisal results and evaluators are mostly internal to the school.

Table 5.3 gives an overview of the involvement of internal and external actors in conducting teacher appraisal processes. While the standards and procedures for teacher appraisal are typically determined at the central level, Table 5.3 shows that the implementation of appraisal processes happens mostly at the school level, i.e. evaluators are typically members of the school leadership team or senior teachers. This is especially the case for regular appraisal for performance management and promotion. Several countries, including Australia, Chile, France, Israel, Luxembourg, Portugal and Sweden involve both internal and external evaluators in teacher appraisal processes.

Table 5.3 Evaluators involved in teacher appraisal processes (2011-12)

	Completion of probation	Performance management			Rewards scheme
		Regular appraisal	Registration	Promotion	
External evaluator(s) such as education authority, central agency, Inspectorate, teacher organisation, accredited evaluator	Ireland (ISCED 1), Slovenia	Mexico	Australia, Sweden	--	Chile, Mexico
School-level evaluator(s) such as school principal, peer evaluator, tutor, school board members	Canada, Ireland (ISCED 2-3), Italy, Netherlands, New Zealand, Slovak Republic, United Kingdom (Northern Ireland)	Australia, Austria (ISCED 1), Belgium (Fl.), Canada, Czech Republic, Hungary, Israel, Korea, Netherlands, New Zealand, Poland, Slovak Republic, Slovenia, United Kingdom (Northern Ireland)	New Zealand	Czech Republic, Estonia, Israel, Korea, Poland	Korea
Both internal and external evaluators are involved	Australia, France, Israel, Luxembourg, Sweden, Portugal	Chile, France, Portugal	--	--	--

Note: In Austria, the state/regional authority may be involved in teacher appraisal processes in ISCED 2-3 in the case of complaints.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Capacity for conducting the appraisal process

Across countries, a range of actors are involved in teacher appraisal. An area to which policy often does not devote enough attention is that of capacity development for appraisal. This is a crucial area to ensure the legitimacy of evaluators and to make an effective use of appraisal results. The effectiveness of appraisal crucially depends on whether evaluators have the knowledge and skills to appraise teachers reliably in relation to established criteria, and also on whether teachers are prepared to use the results of appraisal in ways to improve their performance. Hence, it is important that all those involved in teacher appraisal receive adequate information and professional development opportunities to optimise the outcomes of the process. This includes the teachers themselves as the recipients of appraisal as well as evaluators within the school such as teacher peers and the school leader, and staff of external agencies performing teacher appraisal functions.

Developing skills and competencies for teacher appraisal across the school system takes time and requires a substantial commitment from both education authorities and the main actors involved in teacher appraisal. Considerable time is needed for explanation of teacher appraisal, consensus building among stakeholders about the indicators and norms that make up school or teacher quality, preparing and training of evaluators in terms of methodology, techniques and approaches, as well as providing time and resources for schools and teachers to implement and adapt processes at the school level. This section provides more detail regarding the specific groups involved in conducting appraisal processes and the approaches in place to develop their appraisal capacity. As shown in Table 5.4 (below), the following evaluators are frequently involved in appraisal processes across countries.

Central education authorities

While central education authorities typically play a prominent role in determining procedures for appraisal (see above), they tend to be less frequently involved in the actual appraisal process as evaluators. This may be explained by the difficulty of reviewing the performance of all teachers individually in a centrally organised appraisal process. Hence, central or state education authorities or governments rarely have full responsibility for conducting the teacher appraisal process, but they may share the role of evaluator together with other agencies or the school leader.

For the completion of probation, the central/state education authorities have the full responsibility for conducting the teacher appraisal processes in Slovenia only, and they are involved in the process together with other evaluators in Australia, Israel and Sweden. For regular appraisal in the context of performance management processes, the central or regional authorities have responsibility for conducting the process only in Mexico. In Sweden, a central agency (the National Agency for Education) is in charge of the registration process. For reward schemes, the central education authorities are involved in the appraisal process in Mexico only. In Austria, the state or regional education authorities are involved in appraisal processes only in the case of complaints. In Portugal, the education administration only intervenes with regard to appeals against decisions of the school body that appraises teachers.

Developing teacher appraisal systems can be a highly technical undertaking, and central education authorities and agencies that design teacher appraisal systems have a lot to benefit from evidence on best policies and practices gathered both nationally and internationally. To build central capacity, it is important that different institutions

supporting teachers and the teaching career engage with each other and with the research community more widely so as to ensure that the system is supported by scientific advice and evidence. Box 5.10 provides an example of how central agencies are co-operating with experts in Chile to continuously develop the national teacher appraisal system.

As explained by Isoré (2009), educational researchers and experienced teachers may play a key role in providing advice for the design of appraisal systems. Based on their own experience and research, they can be in a good position to provide expertise on the definition of "good teaching" and the identification of relevant criteria and instruments to appraise teachers (Ingvarson et al., 2007). In addition, conducting a pilot implementation before the full implementation of a new teacher appraisal system can also be helpful in supporting central learning about effective teacher appraisal. It can help policy makers ensure the validity and reliability of a system before full-scale implementation and bring to light potential flaws. It can also allow central system designers to learn directly from the feedback of stakeholders involved in the pilot (Isoré, 2009).

Box 5.10 Developing central expertise for teacher appraisal in Chile

In Chile, teacher appraisal relies on the competencies of several agencies at the central level that co-operate regularly so as to assure the quality of the process. While the Ministry of Education holds the political and management responsibility for teacher appraisal, the technical coordination of the process is exercised by the Centre for Training, Experimentation and Pedagogical Research (CPEIP), which in turn is legally obliged to receive independent scientific advice from universities with expertise in the area. There is recognition at the central level that the implementation of teacher appraisal is a very complex process including a range of both scientific and logistical tasks that could not have been fulfilled effectively by the Ministry or the CPEIP alone.

There is a specific national team responsible for the *Docentemás* teacher appraisal process. This team consists of 36 staff including professionals, technicians and administrative staff, most of them with a background in education or psychology. Under supervision of the CPEIP, the *Docentemás* team collaborates in all aspects of the process, such as the design of assessment instruments and guidelines, logistical aspects of the implementation, selection and training of peer evaluators, evaluation of the teacher portfolios, development and maintenance of information systems and preparation of results reports. The *Docentemás* team uses feedback from teachers who were appraised with the aim of continuously improving the capacity of its own staff.

The close association with the *Docentemás* team ensures that the system is based on scientific advice as well as national and international research evidence. In addition, several universities providing initial teacher education are closely associated to the process. In particular, the portfolio evaluation centres are located within a range of universities across the country. According to CPEIP, involving the universities in the process is essential to make the process legitimate in the eyes of the profession. It also allows building capacity and generating institutional learning within the universities themselves, which may help them align initial teacher education with the expectations of the teacher appraisal process.

Source: Santiago et al. (forthcoming).

School leaders

Given their role as the direct supervisor of teachers in most countries, school leaders typically play a prominent role in teacher appraisal processes. In some instances, they also share appraisal responsibilities with peer evaluators (often senior teachers) and/or external evaluators.

School leaders or managers play a role as evaluator of teachers in the majority of countries and for most appraisal types. For the completion of probation, they are involved as evaluators in all countries for which information is available except Slovenia. School leaders share their responsibility as evaluators with central/state education authorities in Israel and Sweden, with the Inspectorate in France (ISCED 2 and 3) and with a teacher education faculty in Luxembourg. They co-operate with other school-level staff (such as supervisors, peer evaluators, mentor teachers, the school board or school-level evaluation committees) in Australia, France, Italy and the Slovak Republic.

For regular appraisal for performance management, school principals have full responsibility as evaluators in Canada, the Czech Republic, Israel, the Netherlands, Poland, the Slovak Republic and Slovenia. They share this responsibility with external accredited evaluators and a municipal evaluation commission in Chile, with the Inspectorate in France and with teacher peer evaluators in Portugal. Other school-level professionals or the school board are also involved in addition to the principal in Australia, the Flemish Community of Belgium, Hungary and New Zealand. In the Netherlands, the school principal as the competent authority represents the school organising body for both completion of probation and regular appraisal for performance management. In New Zealand, a member of the school leadership team holds this responsibility for registration purposes. For appraisal for promotion, school leaders are responsible for conducting the process in the Czech Republic, Israel and Poland. In Israel, while the school leader makes the final decision for the appraisal for performance management and for promotion, he or she may consult with other school staff regarding teacher appraisal. In Korea, the school principal shares responsibility for appraisals for promotion and for rewards with peer evaluators.

It is difficult to envisage effective teacher appraisal without effective pedagogical school leadership. Education systems have increasingly recognised the importance of such leadership in raising standards, as substantiated in an OECD report (Pont et al., 2008). There are many advantages to having the principal and/or other teachers as the evaluators, especially in developmental appraisal, given their familiarity with the context in which teachers work, their awareness of the school needs and their ability to provide quick and informed feedback to the teacher.

A common challenge in many countries is the lack of well-established pedagogical leadership in schools. In many countries, school leaders have traditionally held a more administrative role and may not have been well trained and prepared to engage in the appraisal of their staff (see also Chapter 7). In systems where school leadership is largely concentrated in one person (the principal) rather than being distributed in a school leadership team, school leaders are also facing challenges to find the time to thoroughly observe and appraise the practice of each teacher. Where hierarchical structures are flat with the school leader being perceived as a *primus inter pares*, such as in Norway, having ambitions for pedagogical leadership including classroom observation may not always be well regarded by teachers and school leaders may be hesitant to exercise such leadership (Nusche et al., 2011a).

In addition, most principals have no prior training in appraisal methods and might not have the content expertise relevant to the teaching areas of the teacher being appraised. They may have limited skills and experience in making adequate observations and judgements about the quality of teaching practice. Especially in contexts where there is little external guidance, schools leaders might be coming to their judgements in isolation with the risk that they might be out-of-line and too limited in their expectations in

comparisons with the standards being applied in teacher appraisal in other schools. In order to address this challenge, it is important for the national level to provide leadership in developing nationally shared standards and criteria for appraisal as well as resources and guidance for school-staff to support effective processes.

There need to be clear criteria for classroom observation and school leaders need to receive training to conduct these adequately. In New Zealand, for example, the New Zealand Teaching Council (NZTC) provides resources and support measures to ensure that principals can undertake effective appraisals and that staff are supported/ guided through the processes (Nusche et al., 2012). In addition, given their crucial pedagogical role, it is important that school leaders benefit from dedicated training and professional development opportunities. Across countries participating in the OECD Review, there were few examples of professional development for school leaders dedicated specifically to methods for teacher appraisal. However, in several countries, a component regarding teacher appraisal was included in broader school leadership development programmes (Box 5.11).

Box 5.11 Supporting school leaders in developing skills for teacher appraisal

In the **Czech Republic**, as part of European Social Fund (ESF) programming, several projects have been developed to improve competencies for teacher appraisal. In particular, the "On the Road to Quality" project, launched in 2009, aims to develop instruments for teacher appraisal as part of school self-evaluation and to build teachers' capacity for understanding and implementing appraisal and evaluation approaches. One aspect of the project was the development of a 360-degree feedback tool for middle management staff in schools. Some of the regional authorities have also developed programmes to strengthen the approaches and methods used for teacher appraisal across schools. For example, in the Moravian-Silesian region, a methodology was developed to support school principals in teacher appraisal, as part of an ESF-funded project ("The Chance").

In **Norway**, where there is little tradition for regular classroom observation by principals, a new national education programme for principals was introduced in 2009. The education programme is initially targeted at newly employed principals who have been in the position for less than two years. It will then be extended for more long-standing principals who have not received such an education. The overall aim of this new initiative is to better equip principals for their role as leaders, and in particular for taking a stronger role in guiding the teaching and learning processes at school. It is expected that as principals are become better prepared for pedagogical leadership, they will also become more confident in appraising and providing feedback to their teaching staff. It is hoped that this will help increase the acceptance among teachers of school leaders observing classrooms and appraising teaching performance. The framework for school leader competences defines four main competences. The appraisal of staff is not included as an explicit competence area, but it is at the core of competence area one: "the pupils' learning results and the learning environment" Under this heading, the competence framework points out that "the head teacher's ability to lead the learning process and guide teachers in this process will be decisive" (Norwegian Directorate for Education and Training, 2008). Among the skills and attitudes principals should be able to master in this area, many relate to appraising and guiding teachers' practices: (1) setting goals for teaching work; (2) setting standards for quality in working processes and being able to enforce these; (3) following up on and giving feedback to individual co-workers; (4) creating pride, aspirations and a desire to achieve results in teachers; (5) guiding and giving feedback to teachers; and (6) challenging teachers and setting definite demands on quality.

Sources: Santiago et al. (2012b); Nusche et al. (2011a).

Teacher peers

In some systems, teacher appraisal is based on reviews by teacher peers. Such peer evaluators are typically "accomplished" current teachers who are recognised as having in-depth subject knowledge and pedagogical expertise, as highly proficient and successful practitioners, able to guide and support others in the teaching process. Peer evaluators may be teachers internal or external to the school of the appraised teacher.

For the completion of probation, peer evaluators and/or supervisors from the same school are involved in the appraisal process in Australia and France. In Italy, an evaluation committee of on average four teachers takes part in the appraisal process together with the school principal. In the Slovak Republic, mentor teachers as well as examination committees nominated by the school principal are in charge of the appraisal process for completion of probation. For regular performance management, peer evaluators are involved in many countries including Australia, Chile, Hungary, Korea, and New Zealand. In Portugal, peer appraisal is the dominant feature of performance management. In Korea, peer evaluators are involved in regular appraisal, appraisal for promotion and appraisal for rewards. For the rewards scheme in Korea (performance-based incentive system), schools are required to set up a screening committee to decide incentive levels and set standards for provision and the school principal makes final decisions regarding incentive provision standards via the deliberation of the committee. In Northern Ireland in the United Kingdom, the school leader may designate a teacher reviewer to conduct appraisal processes under the Performance Review and Staff Development Scheme.

The use of peer appraisal is often more common in developmental appraisal. Research points towards the power of developmental peer observation with clear foci and effective feedback in making a strong contribution to improving the quality of teaching. Peer appraisal can be important both for providing evidence about teachers' practice and for providing support for professional growth (Goldstein, 2007; Goldstein and Noguera, 2006; Milanowski, 2005; Wei et al., 2009), and peers often learn best from each other (Coggshall et al., 2011; Jackson and Bruegmann, 2009).

In the United States, for example, several districts use the Peer Assistant and Review (PAR) programmes for teacher appraisal. In these programmes, expert mentor teachers or coaches support new teachers, as well as experienced teachers who are struggling, and they conduct some aspects of the teachers' appraisal. Based on the appraisal, teachers develop personalised professional development plans directly focused on their identified strengths and weaknesses and work together with mentor teachers to achieve goals outlined in the plan. The programmes also involve a process where a panel of teachers and leaders make recommendations about personnel decisions based on evidence from the appraisal. Several studies argue that inclusion of peers as reviewers and coaches created a more transparent process, and that the approach can contribute to improving teaching while also allowing systems to identify teachers for intensive assistance, continuation, tenure, and in some cases, termination (Goldstein and Noguera, 2006; Klinger et al., 2008; Darling-Hammond et al., 2012).

When peers are the evaluators in accountability-driven teacher appraisal procedures, issues of legitimacy are particularly relevant to address. In such systems, it is particularly important to build capacity in appraisal methods by preparing not only principals but also peer evaluators to undertake specific appraisal functions. Box 5.11 provides examples of initiatives to support peer evaluators in developing their skills for fair and effective peer appraisal (Box 5.12).

Box 5.12 Building capacity for peer appraisal

Teacher appraisal in **Portugal** relies almost entirely on peer appraisal. All key roles in teacher appraisal, including performance appraisal as well as co-ordination, counselling and pedagogical supervision, are exercised by teachers. To enhance capacity for appraisal at the school level, the Ministry of Education entered into a contract with a higher education institution as the managing organisation responsible for launching an in-service training system for teacher appraisal. In the first half of 2011, 50 teachers with a Master's degree in the field of evaluation were identified from Portugal's five educational regions to participate in specialised training on teacher appraisal including classroom observation. In this post-graduate training, particular emphasis was placed on classroom observation, as this was seen as the area that could have the greatest impact on improving teaching and learning. Upon completion of the training, it was expected that this first group of highly qualified teachers would be able to act as multipliers and provide training in teacher appraisal to other the teachers in their schools.

One of the characteristics of **Chile**'s teacher appraisal approach (*Docentemás)* is the high involvement of practising teachers as evaluators. The participation of teachers at various stages of the appraisal process contributes to building ownership and appraisal competency among teachers and may also help them to understand and benefit from their own appraisal to a greater extent. Practising teachers can apply to two key roles in the appraisal process: (1) as evaluators of teacher portfolios in one of the centres set up for this purpose by *Docentemás* in various universities; and (2) as peer evaluators who conduct peer interviews and participate in the municipal evaluation commissions. For both roles, intensive preparation processes have been set up to build the capacity of those selected. The portfolio evaluators are trained in a one-week training session, where they work together with specialists on concrete examples of different performance levels. The training sessions comprise individual and group work in which teachers discuss judgements about proficiency levels. This is followed by a test period where the evaluators apply what they have learned, internalise the portfolio evaluation processes and benefit from group discussion about the results. The peer evaluators are selected and trained by the national *Docentemás* team or the local university in charge of the process. Only teachers who have been previously rated as Outstanding or Proficient can apply to become peer evaluators. They receive training in two full-day seminars, during which they learn about the six questions to be asked in the interview and the rubrics to be applied in assigning performance levels. The training also includes exercises and feedback to the participants. At the end of this training phase, there is another selection process and not all of those initially selected will be retained as peer evaluators.

Sources: Santiago et al. (2012a; Santiago et al. (forthcoming).

Other evaluators

Some countries draw on evaluators that do not readily fit the categories established above. In Australia, the teacher professional organisation is in charge of conducting the appraisal of teachers for registration. In Chile, a municipal evaluation commission composed of trained peer evaluators from the municipality is involved in the process in addition to the school principal and external accredited evaluators.

External inspectors may also play a role. Among the countries participating in the OECD Review, only France and Ireland report that inspectors are routinely involved in teacher appraisal. In France, inspectors play a role in appraisal for the completion of probation (at ISCED levels 2 and 3) as well as in regular appraisal for performance management, and they do so in collaboration with the school principal. Ireland reports that the Inspectorate is involved as evaluator in teacher appraisal for the completion of probation, but only at the primary school level. In addition, in Austria, an inspector from the state or regional education authority may be involved in the case of an appeal or complaint. In the majority of countries, however, inspectors more typically undertake classroom evaluations with a focus on evaluating overall teaching quality in the school, and not for the purpose of individual performance appraisals.

Table 5.4 Evaluators involved in teacher appraisal processes (2011-12)

Country	Probation	Performance management			Rewards
		Regular appraisal	Registration	Promotion	
Australia	State education authorities; school principal; supervisor; peer evaluator at the same school	School principal; member of school leadership other than school principal; supervisor; peer evaluator at the same level	Teacher professional organisation	a	a
Austria	a	School principal; ISCED 2-3: state or regional education authority in case of appeal or complaint	a	a	a
Belgium (Fl.)	a	School principal (1st evaluator); school organising bodies (2nd evaluator)	a	a	a
Belgium (Fr.)	a	a	a	a	a
Canada	School principal	School principal	a	a	a
Chile	a	External accredited evaluators; school principal; teacher from another school; municipal evaluation commission (composed of local peer evaluators)	a	a	Central education authority
Czech Republic	a	School principal	a	School principal	a
Denmark	a	a	a	a	a
Estonia	a	a	a	Special commission	a
Finland	a	a	a	a	a
France	General Inspectorate; school principal; supervisor	General Inspectorate; school principal	a	a	a
Hungary	a	School principal; evaluators specified in the school's quality assurance programme	a	a	a
Iceland	a	a	a	a	a
Ireland	Inspectorate of the Department of Education and Skills (ISCED 1); school leadership (ISCED 2, 3)	a	a	a	a
Israel	School principal; central education authority	School principal	a	School principal	a
Italy	School principal; school-based evaluation committee	a	a	a	a
Korea	a	Peer evaluators at the same school	a	School principal; peer evaluators at the same school	School principal; peer evaluators at the same school
Luxembourg	School principal; teacher education faculty of University of Luxembourg	a	a	a	a
Mexico	a	Central education authority	a	a	Central education authority or government
Netherlands	School principal representing school organising body (competent authority)	School principal representing school organising body (competent authority)	a	a	a

Table 5.4 Evaluators involved in teacher appraisal processes (2011-12) (*continued*)

Country	Probation	Performance management			Rewards
		Regular appraisal	Registration	Promotion	
New Zealand	School principal	Member of school leadership team; peer evaluator from the same school	Member of school leadership team	a	a
Norway	a	a	a	a	a
Poland	a	School principal	a	School principal	a
Portugal	Teachers from the same school; collegiate body within the school (chaired by principal); trained teachers from other schools	Teachers from the same school; collegiate body within the school (chaired by principal); trained teachers from other schools	a	a	a
Slovak Republic	Mentor teacher; examination committee nominated by the school principal	School principal	a	a	a
Slovenia	Central education authority	School principal	a	a	a
Spain	a	a	a	a	a
Sweden	Central education authority; school principal	a	Central agency (National Agency for Education)	a	a
United Kingdom (Northern Ireland)	School principal or a teacher reviewer designated by the school principal	School principal or a teacher reviewer designated by the school principal	a	a	a

Notes: a = not applicable; m = information missing

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Using multiple evaluators

The participation of multiple evaluators is often seen as key to successful teacher appraisal practices. Several researchers in this field recommend that at least more than one person should be involved in judging teacher quality and performance (Peterson, 2000; Stronge and Tucker, 2003). For example, "360-degree evaluation systems", which incorporate the participation of many different evaluators, allow for the integration of several different perspectives in the appraisal process (Danielson and McGreal, 2000). This can help increase the accuracy and fairness of the appraisal process, taking into account the complexity of what a "good" teacher should know and be able to do (Danielson, 1996, 2007; Peterson, 2000).

For example, in a study conducted in Chicago, Sartain et al. (2010) found that school leaders' appraisal of teacher performance tended to reflect greater leniency than that of other evaluators, perhaps because school leaders were more aware of context variables that may impact teacher performance. Some studies also indicate that while principals tend to be accurate in identifying very high and very low performing teachers, they sometimes show less ability to distinguish between teachers in the middle of the distribution and may be influenced by a number of affective or non-performance factors (Bolino and Turnley, 2003; Jacob and Lefgren, 2005, 2008). On the other hand, external evaluators may appraise teachers in relation to frameworks and professional standards and know the specificities of content and skills required for each teaching area, but they are less able to adapt the processes to the school context, challenges and values

(Anderson and Pellicer, 2001). The involvement of more than one evaluator is particularly important in high-stakes appraisal systems.

For formative appraisal, there are also advantages in drawing on the perspectives of several evaluators. Peers and colleagues who have the same characteristics, teach the same subject and/or to the same students are more likely to obtain the confidence of the teacher being appraised. Teachers may therefore more easily engage in self-reflection about their practices, and express their feelings and concerns during interviews, without fearing potential sanctions. Peers can also provide qualitative feedback based on their own experience in the relevant teaching area (Isoré, 2009). But principals are essential to link the teachers' learning needs to the further professional development opportunities in line with broader school goals and human resource needs. They are also more likely to provide informal continuing feedback to the teacher throughout the year and not only during the formal appraisal process. More generally, school leaders are essential to make performance improvement a strategic imperative, and help considering teacher appraisal indispensable to teacher and school broader policies (Heneman et al., 2007; Robinson, 2007; Pont et al., 2008).

In designing teacher appraisal systems drawing on multiple evaluators, it is important to acknowledge that the process may require considerable time and resources of both evaluators and evaluatees. As a consequence, system designers should carefully review the requirements that are made of teachers and designated evaluators. For the appraisal process to have beneficial results, it should allow teachers to have time to reflect on their own practice. As a consequence, it should be considered that both teachers and evaluators are partly released from their duties when undertaking the appraisal process, especially for systems that involve considerable additional workload (Heneman et al., 2006; Isoré, 2009).

Enhancing the skills of teachers to benefit from their own appraisal

There is no internationally comparable information regarding the extent to which teachers are prepared to participate actively in their own appraisal processes and benefit from the results. But anecdotal evidence from the country-specific *OECD Reviews of Evaluation and Assessment in Education* suggests that such preparation is quite limited across countries.

Guaranteeing that teachers are provided with support to understand the appraisal procedures is essential both to engage teachers with the appraisal process and to enhance their performance. For the process to be effective, teachers must know what is expected from them to be recognised as "good" teachers before the process starts. This requires not only complete transparency in the appraisal criteria and procedures but also ensuring that teachers take ownership of the process through support and coaching. In the United States, for example, the *Guide to Understanding National Board Certification* responds to teachers' concerns in relation to the characteristics of the NBPTS appraisal. It explains the system (who is concerned, what the process consists of, how the scores are established, etc.) and gives advice to help teachers succeed (what to include in a portfolio, which exercises to get prepared, examples and ideas from past candidates and trainers) (AFT and NEA, 2008).

Another way to prepare teachers for their appraisal is to link the teacher appraisal system with initial teacher education and professional development. In some countries participating in the OECD Review there were concerns that what teachers learned in their initial preparation was poorly aligned to the standards and criteria of "good teaching" in

teacher appraisal later on. There is a need to ensure that where national standards for the teaching profession exist, these are consistently used in all initial teacher education programmes. Its domains and criteria regarding good teaching should be applied throughout initial teacher education so that teachers already have a clear understanding of what is expected from them when they enter the profession. Self-appraisal, appraisal by others and feedback also should be aspects offered in initial teacher education so that teachers are prepared for such processes.

Induction and mentoring programmes for new teachers can further ease the transition between initial education and school-level appraisal processes. As highlighted by the OECD (2010), recent research indicates that beginning teachers can benefit from mentoring programmes, but it is important to ensure the quality of such programmes. Mentors should be carefully selected, be given adequate time to carry out their tasks and be well prepared for their tasks (Hobson et al., 2009, in OECD, 2010). Developing mentor teachers at the school level can also be a way to distribute school leadership more broadly. Mentors can play a key role in helping new teachers understand existing teaching standards, self-appraise their practice and receive feedback for improvement (Santiago et al., forthcoming).

Use of results

This section is concerned with how teacher appraisal processes are followed up and how results are utilised for further decisions or actions. Examples of mechanisms to use appraisal results include performance feedback, professional development plans, career advancement and financial and other rewards.

Formative use of results

A key objective of teacher appraisal is to identify areas for professional development for individual teachers, leading to the preparation of individual improvement plans which take into account the overall school development plan. Pedagogical leadership at the school level plays a key role in ensuring the effectiveness of this link (Pont et al., 2008). The resources made available for professional and school development are another key element.

Information collected from countries participating in the OECD Review indicates that all types of teacher appraisal (except explicit reward schemes) may potentially influence future professional development activities (Table 5.5). Regular teacher appraisal as part of performance management is most typically connected to professional development activities or plans. It systematically influences professional development in Australia, Korea, Mexico and Northern Ireland in the United Kingdom and it is expected/intended to do so in Austria, the Flemish Community of Belgium, some provinces/territories in Canada, France, Israel, the Netherlands, New Zealand and Slovenia. In Chile, it systematically results in a professional development plan for teachers who have obtained a "basic" or "poor" rating and in Portugal this is the case for teachers who have obtained an "insufficient" rating. In the Czech Republic, Hungary, Poland and the Slovak Republic, the link between regular appraisal for performance management and professional development is not prescribed nationally, but it may well exist at the school level. Practices vary across schools depending on internal regulations.

Table 5.5 Influence of teacher appraisal on professional development (2011-12)

	Completion of probation	Performance management		
		Regular appraisal	Registration	Promotion
It systematically influences professional development activities	United Kingdom (Northern Ireland)	Australia, Chile[2], Korea, Mexico, Portugal[2], United Kingdom (Northern Ireland)	Australia	Czech Republic[3]
It is expected/intended to influence professional development activities	Australia, Canada, Ireland, Israel, Netherlands, New Zealand	Austria, Belgium (Fl.), Canada, France, Israel, Netherlands, New Zealand, Slovenia	New Zealand	Israel
It may influence professional development activities, depending on school policies and practices	Slovak Republic	Czech Republic, Hungary, Poland, Slovak Republic	--	--
It does not influence teacher professional development	France[1], Italy, Luxembourg, Slovenia	--	Sweden	Estonia, Korea, Poland

Notes: (1) France: But a negative appraisal may result in a second year of stage; (2) Chile, Portugal: It systematically results in a professional development plan for teachers who have obtained a low rating only; (3) Czech Republic: It influences professional development if connected with promotion to particular professional status.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Appraisal at completion of probation also influences professional development in Northern Ireland in the United Kingdom, and it is expected to do so in Australia, some provinces/territories in Canada, Ireland, Israel, New Zealand and the Netherlands. Registration systematically influences professional development in Australia and is expected to do so in New Zealand. The promotion scheme in Israel is also expected to inform future professional development. In the Czech Republic, appraisal for promotion influences professional development if it is related to accession to a particular professional status, such as that of pedagogical advisor.

Providing feedback for individual professional development

A logical chain between the performance appraisal and continuing professional development opportunities is essential to improve teaching practice (Ofsted, 2006). The identification of individual teachers' strengths and weaknesses is important to choose from a wide range of possible professional development activities the ones that meet individual teachers' own needs against each of the priorities in the school improvement plan. Seeing the appraisal procedures as a basis for future practice improvement is critical to implement a system in which every single teacher will feel concerned by the appraisal and the relevant professional growth opportunities, regardless of their current level of performance (Isoré, 2009).

However, among the teachers participating in TALIS, over 40% reported that they did not receive suggestions for improving aspects of their work and 44% agreed with the statement that teachers' work was reviewed merely to fulfil an administrative requirement. Also, according to the reports of principals in TALIS, only 56.6% of teachers were in schools where the identification of a specific weakness in teacher appraisal leads always or most of the time to establishing a professional development plan for the teacher. These are worrying results. Without a clear link to professional growth opportunities, the impact of teacher appraisal on teaching and learning will be relatively limited (Goe et al., 2012). As a result, the appraisal process may not be taken seriously or encounter mistrust or apathy by the teachers being appraised (Danielson, 2001; Milanowski and Kimball, 2003; Margo et al., 2008). Ideally, teacher appraisal should

allow teachers to receive tailored feedback and such feedback should be followed with opportunities for continuous learning in identified areas through professional development, mentoring and other means (Hill and Herlihy, 2011).

There is also a need to envisage teachers' learning as something broader than participation in in-service training courses. According to Timperley (2011), the term "professional development" is now often associated with the delivery of some kind of information to teachers in order to influence their practice, whereas "professional learning" refers to a more internal process in which teachers create professional knowledge through interaction with this information in a way that challenges previous assumptions and creates new meanings. Such professional learning cultures need to be supported and sustained by effective pedagogical leadership providing adequate levels of challenge and support to teachers. All teachers, including the highly effective ones, need opportunities to learn and grow in the teaching profession (Randi and Zeichner, 2004). Box 5.13 provides examples regarding the use of appraisal results for professional development from Korea and Memphis, Tennessee in the United States.

Box 5.13 Linking teacher appraisal to professional development

In **Korea**'s Teacher Appraisal for Professional Development, once appraisal procedures are completed, evaluation sheets are collected and drafted into a final report for each teacher. Results of the peer review process are written by up "appraisal management committees" at each school. Upon receiving their appraisal results report, teachers prepare their own "plans for professional development" (including training attendance plans) and submits these to the appraisal management committee. The committee brings together the professional development plans and the appraisal results of all appraised teachers, and drafts a "synthetic report on Teacher Appraisal for Professional Development" to submit to the principal and vice-principal. The synthetic report must include: implementation plan and progress of appraisal; overall appraisal results (excluding results for individual teachers); general features of appraisal (appraisal provided by parents, students and peer teachers; strengths and weaknesses of the school's teachers as revealed by appraisal); teachers' demand for training including autonomous in-service training; fields of training requested by the teachers; the school's next-year plans to provide consulting and training programmes for teachers' professional development; budget estimation; proposals and requests to the local education authority (demands for the establishment of new training programmes, requests for budget support for in-service training by the education office, etc.).

Based on the appraisal results, local education authorities provide excellent teachers with a "study and research year" (similar to sabbatical years given to university faculty) as a way of granting opportunities for teachers to further build their professional capacity. Underperforming teachers are obliged to undertake short- to long-term training programmes depending on their appraisal results. Also, regardless of appraisal outcomes, local education offices support teachers with customised training programs, so as to foster an atmosphere of self-study and self-improvement among teachers. Individual appraisal reports are notified only to the concerned teacher and principal. Aside from that, according to the Act on the Protection of Personal Information Contained by Public Institutions, no appraisal results are disclosed to others. For students and parents, only the results tendency of all teachers appraised in a school is provided.

Memphis, Tennessee in the United States has developed a system that explicitly links professional learning to teacher appraisal. In Memphis City Schools, appraisal is based on teaching standards, and professional development is linked to teachers' competence on the standards. Thus, a teacher who has poor performance on a specific indicator on a teaching standard can find professional growth opportunities related to that indicator. Memphis City Schools publishes a professional development guide each year that lists the professional growth offerings by standard and indicator. In addition, most of the professional development courses are taught by Memphis City School teachers, ensuring that the course offerings will be relevant to the contexts in which these teachers work

Sources: Kim et al. (2010); Memphis City Schools, *www.mcsk12.net*; Santiago et al. (forthcoming).

Providing feedback to initial teacher training institutions

To enhance teacher learning in the long run, appraisal results could also be used to improve initial teacher education programmes. Teacher education institutions in most countries are poorly informed about how their graduates perform in schools. However, teacher appraisal systems can provide very important information about the skills needs of teachers and it is important that such information be made easily accessible to teacher education institutions. This is an area where electronic data management systems could play a critical role.

In some countries, including the United States, there is a push towards ensuring that teacher education programmes: *(i)* receive feedback about the performance of their graduates; and *(ii)* are held accountable for improving their instruction in order to ensure better performance of teachers for local contexts (National Research Council, 2010). For example, federal policies in the United States require teacher education programmes to document the efforts they have made to gather information about the satisfaction of local school districts with the teachers prepared in their institutions. In addition, several states in the United States collect teacher evaluation results specific to each teacher education programme and determine whether the programmes' graduates are performing adequately. Programmes whose graduates are not performing effectively must devise a plan to improve their effectiveness – through more selective admissions, better education, or a combination of both (Santiago et al., forthcoming).

Summative use of results

Appraisal of teacher performance can also be used to determine career advancement, award performance rewards or establish sanctions for underperforming teachers. It constitutes an opportunity to recognise and reward teaching competence and performance, which is essential to retain effective teachers in schools as well as to make teaching an attractive career choice (OECD, 2005).

Career decisions

First of all, appraisal results may be used to influence career decisions. Most countries do not directly link teacher appraisal results to teacher pay but, instead, to career progression (therefore establishing an indirect link with salaries). As shown in Table 5.6 below, most teacher appraisal models relate results to the speed at which the teacher progresses in the career. In addition, teacher appraisal can also be used to make decisions at key points in a teacher's career. For example, the results of teacher appraisal can be used to make tenure decisions at the end of the probationary period, contract renewal and registration stages, and might influence the chances of an application to a given school post. In the Flemish Community of Belgium, teacher appraisal does not influence career decisions or career advancement. In the other education systems that reported having formal frameworks for teacher appraisal, there appears to be at least one type of appraisal that influences career decisions or advancement.

Logically, all countries that have appraisal processes explicitly designed for promotion purposes use the appraisal results for this purpose. However, it is important to note that many countries that do not have specific processes for teacher appraisal for promotion use the results of other appraisal processes to inform decisions about a teacher's career advancement. Appraisal for the completion of probation most typically influences decisions about access to a permanent position or fixed term contract. This is the case in Australia, Israel, Italy, Luxembourg and Sweden. It influences career progression in some provinces/territories in Canada, France, Ireland, the Slovak Republic

and Northern Ireland in the United Kingdom, and it may influence career decisions (depending on school policies) in the Netherlands. Regular appraisal for performance management influences decisions about access to a permanent position or fixed term contract in Mexico and contract renewal for non-permanent teachers in Portugal. It influences decisions about career progression or promotion in Australia, France, Hungary, Portugal (for permanent teachers), Slovenia and Northern Ireland in the United Kingdom. It may influence decisions about career advancement in the Netherlands, New Zealand and the Slovak Republic. Registration processes may also influence decisions about the teachers' position or contract type. In Australia, it determines eligibility to seek employment as a teacher. In New Zealand and Sweden, appraisal for registration is a necessary step towards progression to the status of "registered teacher".

Table 5.6 Influence of teacher appraisal on career decisions (2011-12)

Appraisal may influence:	Completion of probation	Performance management			Reward scheme
		Regular appraisal	Registration	Promotion	
Decisions about access to a permanent position or a fixed-term contract	Australia, Israel, Italy, Luxembourg[1], New Zealand[2], Sweden	Mexico, Portugal (non-permanent teachers)	Australia, New Zealand, Sweden	--	--
Decisions about the speed of career progression and/or about promotions	Canada, France, Ireland (ISCED 1), Slovak Republic, United Kingdom (Northern Ireland)	Australia, France, Hungary, Portugal (permanent teachers), Slovenia, United Kingdom (Northern Ireland)	--	Czech Republic, Estonia, Israel, Korea, Poland	--
Decisions about career advancement depending on local/school policy	Netherlands	Netherlands, New Zealand, Slovak Republic	--	--	--
It does not influence career advancement	Slovenia	Austria, Belgium (Fl.), Chile, Israel	--	--	Chile, Korea, Mexico

Notes: (1) Luxembourg: It influences decisions about the terms of the contract; (2) New Zealand: Teachers need to pass probation and achieve registration to continue in the teaching career.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

In many countries participating in the OECD Review, there was concern about the absence of career opportunities for effective teachers. Typically, in most countries, there is no clearly designed career structure for teachers and there are few opportunities for teachers to take on greater responsibilities, receive recognition for high performance or be promoted. The organisational structure in schools in many OECD countries is typically flat with few promoted posts and few explicit means of giving teachers significant lead responsibilities. This is likely to undermine potentially powerful links between teacher appraisal, professional development and career development.

Findings from TALIS confirm a weak link between teacher appraisal and career advancement. Across TALIS countries, only 16.2% of teachers indicated that the appraisal and/or feedback they received led to a moderate or large change in the likelihood of their career advancement. Also, only 26.7% reported that it led to changes in work responsibilities that made their job more attractive (OECD, 2009).

Some countries link teacher assessments with opportunities for vertical promotions to school leadership positions. But the practice of linking outstanding teacher performance to promotions for school leadership positions may not respond well to the needs of most teachers, for two main reasons. First, a good teacher is not necessarily a good manager or leader and the skills required for teaching a classroom and managing a school are not the same. Second, this practice may have adverse effects on teaching quality within a school

because, paradoxically, the best teachers are rewarded by being removed from classroom teaching. To resolve this dilemma, some education systems have attempted to build career options for excellent teachers who wish to remain in the classroom. Box 5.14 provides examples from Australia.

Box 5.14 Advanced Skills Teaching positions in Australia

Teachers in Australia undergo appraisal, on a voluntary basis, to gain promotion positions in schools in recognition of quality teaching performance by applying for Advanced Skills Teaching positions (ASTs). These positions are linked to higher pay and are generally associated with further responsibilities and specific roles in schools. In most cases, teachers do not have to be at the top of the salary scale to apply for these positions which entails a thorough assessment of their performance. Advanced Skills Teaching positions, which exist in almost all educational jurisdictions, for the most part accomplish two important functions: the recognition of advanced teaching skills with a formal position and additional pay; and a better match between teachers' skills and the roles and responsibilities needed in schools through competitions to gain the positions. These have the benefit of rewarding teachers who choose to remain in the classroom rather than to move into management positions.

AST positions embody two key concepts in the teaching profession in Australia. First, they recognise the need to introduce career diversification as a result of the greater variety of roles in schools – e.g. departmental head, team leader, and manager of curriculum development and/or personnel development. Second, they reflect the need to reward teachers for their developing skills, performance and responsibilities, in what constitutes a competency-based professional career ladder. Teachers, as they access AST positions, are expected to have deeper levels of knowledge, demonstrate more sophisticated and effective teaching, take on responsibility for co-curricular aspects of the school, assist colleagues and so on. Access to AST positions involves formal appraisal processes which are more summative in nature.

- **New South Wales** introduced the Highly Accomplished Teacher (HAT) position in July 2009. The HAT position is an initiative of the Smarter Schools National Partnership on Improving Teacher Quality. A HAT is an excellent teacher who models high-quality teaching for his/her colleagues across the school and leads other teachers in the development and refinement of their teaching practice to improve student learning outcomes. HAT positions are classroom-based positions with a reduced teaching allocation to enable them to mentor other teachers, including student teachers, beginning and more experienced teachers, work with university partners and take a role in the school's leadership team. HATs are appointed through a merit selection process which requires, as a prerequisite, application to the NSW Institute of Teachers for consideration of accreditation at Professional Accomplishment or Professional Leadership. These positions are two-year appointments and are limited to 100 positions over the life of the National Partnerships.
- The **Northern Territory**'s Accomplished Teacher status requires applicants to participate in an "inquiry process" over 12 months, based on the Northern Territory Teacher Registration Board Accomplished Standards of Professional Practice for Teaching. The assessment of performance is undertaken by assessment panels and moderation committees and includes the appraisal of teaching modelling and role in curriculum and professional learning. This process was being reviewed in 2011.
- In **Tasmania**, the Advanced Skills Teacher position recognises outstanding classroom teachers and leading staff members. It is targeted at teachers recognised as exemplary practitioners, who are accorded additional responsibilities within their school. It is a promotion available to any permanent teacher who satisfies the application process, operating in a similar way to a salary increment. Positions are advertised by individual schools on a needs basis.
- The **Victorian** school system includes one promotional appointment for those teachers who want to remain in the classroom: Leading Teacher. The programme is intended to serve the dual purpose of recognising outstanding classroom teachers; and providing schools with a human resource to lead various in-school programmes and projects. Schools advertise for Leading Teacher positions on a needs basis – the position is usually associated with a specific anticipated responsibility. The Victorian Department of Education and Early Childhood Development aims to maintain a Leading Teacher profile of 10 to 15% of full-time teaching staff.

Source: Santiago et al. (2011).

Singapore is another example of a country that focuses a lot of attention on the development and diversification of teachers' careers. The country implemented a performance management approach called Enhanced Performance Management System (EPMS) in 2005. EPMS forms part of a broader career and recognition system, the Education Service Professional development and Career Plan (Edu-Pac) designed to help teachers realise their full potential. This structure has three components: a career path, recognition through monetary rewards, and an appraisal system (Lee et al., 2010). For more information, see Box 5.15.

Box 5.15 Singapore: Linking teacher appraisal to career pathways

The Education Service Professional Development and Career Plan (Edu-Pac) in Singapore recognises that teachers have different interests and aspirations and provides three different career tracks for teachers:

- The **Teaching Track** allows teachers to remain in the classroom and advance to the levels of Senior Teacher, Lead Teacher or Master Teacher. This provides an opportunity for teachers to focus on classroom teaching while obtaining a leadership role along with a senior-level salary.
- The **Leadership Track** provides opportunity for teachers to take on leadership positions within the school or at the Ministry of Education.
- The **Senior Specialist Track** allows teachers to join the Ministry of Education's headquarters and as specialists with particular expertise in specific aspects of education.

The Enhanced Performance Management System (EPMS) serves to support teachers' professional and career development and its results inform promotion decisions as part of Edu-Pac. The EPMS process involves performance planning, performance coaching and performance appraisal. Performance planning involves a teacher self-appraisal and a discussion with the teachers' reporting officer (typically a Head of Department) about target setting and performance benchmarking. Performance coaching is ongoing and includes a formal mid-year review between the teacher and the reporting officer. Finally, the performance appraisal at the end of the year includes an appraisal interview and a rating of actual performance against planned performance. Teachers are appraised based on actual achievement as well as potential for future performance. Decisions on the teacher's "current estimated potential" are made in consultation with senior colleagues of the teacher based on observation, dialogue, portfolio evidence and the teacher's contributions to the school and its environment. The final performance grade affects the annual performance bonus received for the year's work as well as promotions to the next level of the career pathway.

Sources: Lee and Tan (2010); Weinstein and Struthers (2012).

Salary increases and performance rewards

In some cases, teacher appraisal influences a teachers' base salary or salary progression. This impact is sometimes difficult to disentangle as salary progression tends to be connected to career progression (see above). In addition, teacher appraisal results might also be used to award rewards to teachers. Typical rewards include: the award of a one-off monetary prize (bonus pay); time allowances and sabbatical periods; opportunities for school-based research; public recognition or awards; changes in work responsibilities; support for post-graduate study; or opportunities for in-service education.

In some instances the focus of the rewards is on group recognition and rewards are at the school or year level rather than individual level.

Obviously, in countries that have specific rewards schemes in place, the results of these processes influence salary levels and distribution of pay allowances. But as shown in Table 5.7, other appraisal approaches may also be connected to rewards.

Appraisal for the completion of probation is rarely connected to monetary or non-monetary rewards. It has an impact on salary levels in France and an impact on opportunities for professional development and innovative tasks in the Netherlands. Depending on school policies, it may also have an impact on salary levels in the Netherlands and in the Slovak Republic (only to the extent that it can allow teachers to progress on the career scale).

Regular appraisal for performance management appears more likely to influence salary levels or pay allowances across countries. This is the case in Chile, the Czech Republic, Hungary, Portugal and Slovenia. In Northern Ireland in the United Kingdom, the Performance Review and Staff Development statement is part of the body of evidence used to inform decisions on pay progression. It may also impact salaries in the Netherlands and Poland depending on school policies. In Australia, France, New Zealand and Portugal, it only impacts on salaries to the extent that it makes teachers progress on the salary scale. Performance management impacts on public recognition or awards in the Czech Republic and on work responsibilities or opportunities for professional development in the Flemish Community of Belgium (depending on school policies), the Czech Republic (depending on school policies), France and the Netherlands.

Appraisal for registration is the least likely to be connected to rewards. Appraisal for promotion is connected to a pay allowance in the Czech Republic and influences salary levels in Poland to the extent that it makes the teacher progress on the salary scale. It also results in public recognition in Estonia, and opens further opportunities for professional development or post-graduate study in the Czech Republic.

Table 5.7 Rewards connected to teacher appraisal (2011-12)

Appraisal may impact on:	Completion of probation	Performance management			Reward scheme
		Regular appraisal	Registration	Promotion	
Salary levels or pay allowance	France[1], Netherlands[2], Slovak Republic[1,2], United Kingdom (Northern Ireland)	Australia[1], Chile[3], Czech Republic[2], France[1], Hungary, Netherlands[2], New Zealand[1], Poland[2], Portugal[1], Slovenia, United Kingdom (Northern Ireland)	New Zealand[1]	Czech Republic, Poland[1]	Chile, Korea, Mexico
Public recognition or award	--	Czech Republic	--	Estonia	--
Work responsibilities, professional development/ study opportunities	Netherlands	Belgium (Fl.)[2], Czech Republic[2], France, Netherlands	--	Czech Republic	--
No impact on salary or rewards	Australia, Canada, Ireland, Israel, Italy, Luxembourg, New Zealand, Slovenia, Sweden	Austria, Canada, Israel, Korea, Mexico	Australia, Sweden	Israel, Korea	--

Notes: (1) Appraisal results only impact on pay levels to the extent that they make the teacher progress in the career and salary scale; (2) Depends on local and/or school policy; (3) Teachers with satisfactory appraisal results may take an additional appraisal which may give them access to salary increases between 5% and 25% of the base salary.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Rewards connected to teacher appraisal may constitute an opportunity to recognise and remunerate teaching competence and performance. However, it needs to be kept in mind that the issues surrounding the relationship between teacher performance and rewards are controversial and potentially divisive in all countries; and research in this field is difficult and has produced mixed results (Isoré, 2009). There seems to be agreement that the design and implementation of performance-based rewards are crucial to their success. As explained in Harvey-Beavis (2003), there is relatively wide consensus that many attempts at introducing performance-based reward programmes in the past were poorly designed and implemented (Mohrman et al., 1996; Ramirez, 2001). Problems in developing fair and reliable indicators, and the training of evaluators to fairly apply these indicators have often undermined attempts to implement programmes (Storey, 2000). One problem identified is poor goal clarity because of a large number of criteria, which restricts teachers' understanding of the programme and makes implementation difficult (Richardson, 1999). Explanations of how, and on what criteria, teachers are assessed may be difficult to articulate. When this occurs, it is almost impossible to give constructive feedback and maintain teacher support for the programme (Chamberlin et al., 2002).

Sanctions for underperformance

Some teacher appraisal systems include the possibility of sanctions for ineffective teachers beyond the standard consequences for career progression. For example, if underperformance persists following a number of appraisals, sanctions might include the removal from teaching duties (for teachers with civil servant status, this might translate into other functions within the school or another career within the civil service), or simply the termination of the contract. But early identification of underperformance is typically accompanied by a plan for in-service training for the improvement of practice.

As can be seen from Table 5.8 below, many countries have a range of potential responses to underperformance. Underperformance at the end of probation typically results in failure to pass the probation, extension of the probationary period or the termination of the teacher's employment. In Australia, some provinces/territories in Canada, Ireland (ISCED 1), Luxembourg, the Slovak Republic and Sweden, it may also have an impact on the future contract, career or salary. In Australia, Ireland (ISCED 1), Israel and Luxembourg it may lead to further examination, appraisal or compulsory training. In New Zealand, provisionally registered teachers who are not successful in the appraisal can take more time and try again.

Underperformance in regular appraisal for performance management most frequently leads to compulsory training and further appraisal. In many countries, it may also have an impact on contract, career advancement or salary levels. In Australia, Austria (in extreme cases only), some provinces/territories in Canada (at the discretion of the evaluator), Chile and New Zealand, there is the possibility for underperformance to lead to transfer, suspension or dismissal of the teacher. In Northern Ireland in the United Kingdom, if a teacher's performance is unsatisfactory in the Performance Review and Staff Development Scheme, there is an informal stage where a programme of support and development is provided. This may be followed by a formal stage which includes the issue of formal written notice, a targeted support programme and ultimately dismissal if a satisfactory standard of work is not achieved.

Underperformance in registration processes may have an impact on contract, career advancement or salary levels in Australia, New Zealand and Sweden, and lead to further appraisal or compulsory training in Australia. Underperformance in appraisal for

promotion purposes can have an impact on contract, career or salary in the Czech Republic, Israel and Poland. In Estonia and Israel, it also leads to further appraisal and/or compulsory training.

Table 5.8 Responses to underperformance of teachers (2011-12)

	Completion of probation	Performance management		
		Regular appraisal	Registration	Promotion
Failure to pass probation/extension of probation	Australia, Canada, France, Ireland, Israel, Italy, Luxembourg, New Zealand, Slovak Republic, Slovenia, Sweden	--	--	--
Impact on contract, career and/or salary[1]	Australia, Canada, Ireland (ISCED 1) , Luxembourg, Slovak Republic, Sweden	Australia, Canada[3], France, Israel, New Zealand, Portugal, Slovak Republic	Australia, New Zealand, Sweden	Czech Republic, Israel, Poland
Further appraisal, compulsory training	Australia, Ireland (ISCED 1), Israel, Luxembourg, New Zealand, United Kingdom (Northern Ireland)	Australia, Belgium (Fl.), Chile, France, Israel, Korea, New Zealand, Portugal, Slovenia, United Kingdom (Northern Ireland)	Australia	Estonia, Israel
Transfer, suspension, dismissal	Australia, France, Ireland (ISCED 2-3)[2], Israel, Italy, Netherlands, United Kingdom (Northern Ireland)	Australia, Austria[2], Canada[3], Chile, New Zealand, United Kingdom (Northern Ireland)	--	--
Depends on local/ school policy	--	Czech Republic, Hungary, Netherlands	--	--
No response	--	Mexico, Poland	--	Korea

Notes: (1) Impact on contract, career and/or salary includes: permanent contract not granted, salary increment withheld, deferral of promotion, registered teacher status not granted or withdrawn; (2) Applied only rarely/in exceptional cases; (3) At the discretion of the evaluator.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Providing a mechanism to identify weaknesses and ensure that underperformance is adequately addressed is an important feature of teacher appraisal. However, results from TALIS indicate that the use of teacher appraisal to address underperformance is not widespread. On average, in most TALIS countries principals frequently reported the outcome of a teacher appraisal that identifies weaknesses to the teacher concerned and engaged in discussions on how to remedy the weaknesses. However it was much less frequent for the principal to report underperformance to an outside body: 51.0% of principals across TALIS countries indicated that they would never report a teachers' underperformance to another body to take action. TALIS data also shows that a substantial number of teachers across countries had the perception that sustained underperformance is not necessarily addressed: only 23.1% of teachers agreed or strongly agreed that the school principal in their school would take steps to alter the monetary rewards of a persistently underperforming teacher and only 27.9% of teachers agreed or strongly agreed that in their school teachers would be dismissed because of sustained poor performance.

To ensure that all students are taught by capable teachers, it is important that teacher appraisal provides an opportunity to identify incompetent teachers and address the concerns in their performance. However, it has been argued that the identification of underperformance should not be the primary purpose of regular teacher appraisal. This is because the overall impact of an appraisal focussing primarily on the identification of underperformance may be the creation of tensions and fear among teachers, which may

jeopardise the formative function of the appraisal and inhibit teachers' creativity and motivation (Klinger et al., 2008). At the same time, if underperformance is noticed as part of regular appraisal for performance management, weaknesses need to be picked and be adequately addressed.

Pointers for future policy development

This chapter reviewed country approaches to appraising the performance of individual teachers in light of available research and evidence. The policy suggestions that follow are drawn from the experiences reported in the Country Background Reports, the analyses of external review teams in OECD Country Reviews and the available research literature. It should be stressed that there is no single model or global best practice of teacher appraisal. The development of practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing teacher appraisal policies in different countries. In general, there is a need for further research into the impact of different policy approaches to teacher appraisal. The existing evidence base is dominated by research in a few systems with long-established policies on teacher appraisal. As more systems adopt and implement different teacher appraisal policies, there will be a need to collect evidence on how these impact student learning and educational experiences.

Governance

Clarify the purposes of teacher appraisal and ensure that it fits national education objectives

Designing and governing a framework for teacher appraisal entails a range of aspects. First, the purposes of the teacher appraisal framework need to be clearly defined. In particular, it needs to be clear what aspects teacher appraisal seeks to monitor and improve. Second, teacher appraisal needs to be framed in the context of the overall objectives for schooling and the approach to its development should depend on a range of established practices in the school system such as the extent of school autonomy, the existence of national curricula and standards and the culture of evaluation. Third, there needs to be a reflection on the way teacher appraisal is articulated with the remaining components of the evaluation and assessment framework such as school evaluation, student assessment and system level evaluation. Finally, a coherent teacher appraisal system needs to be embedded with well-aligned policies for teacher education, induction and mentoring, support structures, and professional learning opportunities. All of these teacher policy elements should be organised around a shared understanding of what constitutes good teaching.

Resolve tensions between the developmental and accountability functions of teacher appraisal

There are risks that the developmental function of teacher appraisal is undermined when it is too closely associated with high-stakes appraisal for accountability purposes. Therefore, it is not advisable to design appraisal approaches that aim to fulfil the

developmental function of teacher appraisal through a high-stakes accountability-oriented process. Developmental teacher appraisal is likely to benefit from conditions such as a non-threatening appraisal context, a culture of ongoing observation, mutual appraisal and feedback within the school, clear individual and collective objectives, simple school-based appraisal tools and supportive school leadership. In turn, teacher appraisal for accountability is likely to be enhanced by conditions such as central standards and criteria, an independent appraisal component external to the school, a more formal and standardised process and possibilities for teachers to appeal where there are doubts about the fairness of the process.

Establish a coherent framework for teacher appraisal

To build a systematic and coherent system of teacher appraisal, it is important that the approaches to appraisal are adapted to the different stages of a teachers' career and in line with the purposes they are aiming to achieve. A coherent appraisal framework needs to be based on expertise regarding effective design of appraisal systems, draw on multiple appraisal instruments, be conducted by well-trained evaluators and offer differentiated appraisal approaches for teachers at various career stages (Danielson, 2001, 2011; Darling-Hammond, 2012). Developing such a teacher appraisal system may be costly and challenging to implement, but it is critical to reconcile the demands for educational quality, the enhancement of teaching practices through professional development, and the recognition of teachers' knowledge, skills and competencies.

To achieve coherence in appraisal across teachers' career, countries should consider creating a continuum of appraisal approaches linked to professional learning and career advancement. This could start with appraisal at the end of a probationary period, be enhanced by ongoing formative and school-based appraisal and be complemented by periodic summative appraisal for career-progression and accountability purposes. This three-tiered approach will be explored in more detail below.

Establish a mandatory probationary period for new teachers

A formal probationary period for new teachers should constitute the first step in the teaching career. Such a probationary period can provide an opportunity for both new teachers and their employers to assess whether teaching is the right career for them. The satisfactory completion of a probationary period of one to two years teaching should be mandatory before awarding teacher registration or a permanent teaching post. Beginning teachers should be given every opportunity to work in a stable and well-supported school environment, and the decision about completion of probation should be taken by a panel which is well trained and resourced for appraising new teachers. The successful completion of probation should be acknowledged as a major step in the teaching career, corresponding to the access to the first stage of the career.

Consolidate regular developmental appraisal at the school level

In all countries, there needs to be a strong emphasis on teacher appraisal for developmental purposes (i.e. formative appraisal). Given that there are risks that the developmental function is hampered when is too closely associated with a high-stakes teacher appraisal process, a component predominantly dedicated to developmental appraisal, fully internal to the school, should be consolidated. It would be an internal process carried out by line managers, senior peers and the school principal (or members of the management group).

The main outcome would be feedback on teaching performance and overall contribution to the school which would lead to a plan for professional development. It can be low-key and low-cost, and include self-appraisal, peer appraisal, classroom observation, and structured conversations and regular feedback by the school principal and experienced peers. To be effective, appraisal for improvement requires a culture in which there is developmental classroom observation, professional feedback, peer discussion and coaching opportunities. Feedback and discussions should be focussed on the quality of teaching and contributing to the personal and professional growth of those who participate. The results of such processes also allow schools to adapt their professional development offers to the needs of individual teachers and the school goals as a whole.

In order to guarantee the systematic and coherent application of such developmental teacher appraisal across schools, it would be important to undertake the external validation of the respective school processes. An option is that school review or inspection processes, in their monitoring of the quality of teaching and learning, include the evaluation of the processes in place to organise developmental appraisal, holding the school principal accountable as necessary.

Establish periodic career-progression appraisal involving external evaluators

In addition to regular developmental appraisal, teachers could benefit from more summative teacher appraisal at key stages in the teaching career to formalise the principle of advancement on merit associated with career opportunities for effective teachers. Such appraisal would have as its main purposes holding teachers accountable for their practice, determining their career advancement, and informing their professional development plans. This approach would convey the message that reaching high standards of performance is the main road to career advancement in the profession. This could be achieved through a registration system. For fully registered teachers, access to promotions could be organised through a voluntary application process and teachers should be required to periodically maintain their registration status when not applying for a promotion.

Such appraisal, which has stakes for individual teachers, needs to have a stronger component external to the school and more formal processes to ensure fairness across schools. The appraisal should allow making comparisons on the basis of which consequential decisions concerning the teacher can be made. It can be a mostly school-based process led by the school leader (or another member of the management group) but it should include an element of externality such as an accredited external evaluator, for example a teacher from another school with expertise in the same area as the teacher being appraised.

Create a teacher career structure with distinct pathways

In a number of countries the absence of career opportunities for effective teachers undermines the role of teacher appraisal. Schools and teachers are likely to benefit from a career structure for teachers, which should comprise a number of key stages. The different stages in the career should be associated with distinct roles and responsibilities in schools associated with given levels of expertise. Access to each of the key stages could be associated with formal processes of appraisal through a system of teacher registration.

The career structure for teachers should match the different types and levels of expertise reflected in teaching standards. Such alignment would reflect the principle of rewarding teachers for accomplishing higher levels of expertise through career advancement and would strengthen the linkages between roles and responsibilities in schools (as reflected in career structures) and the levels of expertise needed to perform them (as reflected in teaching standards).

This would strengthen the incentive for teachers to improve their competencies, and reinforce the matching between teachers' levels of competence and the roles which need to be performed in schools to improve student learning. It is important that a career structure for teachers increases flexibility to support, recognise and reward excellence in teaching. It should also give possibilities for teachers to have diverse career pathways where some will move more into leadership roles while others choose to remain predominantly teaching in the classroom.

Address the challenges of implementation

It is essential to anticipate the challenges of implementation of teacher appraisal. This includes reconciling the diverging interests of stakeholders, carefully analysing policy alternatives and their likely impact and discussing them with stakeholders to aim towards consensus. Adverse effects are particularly prone to occur when consensus has not been reached on the objectives for teacher appraisal, its importance for the performance of the school system, and the practical options for implementation. Consensus is all the more precious to reach since local and school-level actors are probably in the best position to foresee unintended consequences or judge what is feasible in practice, and since the effectiveness of teacher appraisal heavily depends on their co-operation.

Teacher appraisal and the resulting feedback, reflection and professional development will only work if teachers make it work. To a great extent it is the motivated teacher who ensures the successful implementation of reforms in schools. Hence, it is imperative to find ways for teachers to identify with the goals and values of teacher appraisal arrangements and practices and to encourage their active participation in the process. The expectation is that teachers' engaging in reflective practice, studying their own teaching methods and sharing experience with their peers in schools become routine parts of professional life. In addition, there needs to be a clear understanding of the responsibilities of different actors within the teacher appraisal framework. Education authorities at several levels, agencies in charge of quality assurance, inspectorates, schools, parents, teachers and students play different roles in ensuring improvement and accountability in the teaching profession.

Procedures

Establish teaching standards to guide teacher appraisal and professional development

The key element in a teacher appraisal system is a teacher professional profile (or teaching standards), i.e. a clear and concise statement of what teachers are expected to know and be able to do. These could cover areas as articulated in the Danielson's *Framework for Teaching* (1996, 2007): planning and preparation; the classroom environment; instruction and professional responsibilities. In recognition of the variety of tasks and responsibilities in today's schools and the teaching expertise developed while

on the job, teaching standards could express different levels of performance such as competent teacher, established teacher, and accomplished/expert teacher.

Teaching standards should reflect teachers' tasks in schools and the knowledge and skills that they need to acquire to be effective at the different stages of their careers to achieve student learning objectives. They need to reflect the sophistication and complexity of what effective teachers are expected to know and be able to do; be informed by research; and benefit from the ownership and responsibility of the teaching profession. It also needs to be ensured that teaching standards provide a common basis to organise the key elements of the teaching profession such as initial teacher education, teacher registration, teachers' professional development, career advancement and, of course, teacher appraisal.

It is important to ensure that teaching standards are well understood by all those involved in teacher appraisal processes. There is no justification for anyone, certainly not teachers, to be in the dark as to what is expected of them in their practice, and how that practice will be evaluated throughout their careers. It is equally important that different levels of performance and appraisal criteria are well understood by teachers. Various tools can contribute to this understanding, such as short video clips illustrating high and low performance or sample portfolio entries with an assessor's comments. This would allow teachers to reflect on their own practices and engage in professional learning.

Use multiple instruments and sources of evidence

Establishing effective teacher appraisal procedures is challenging at several levels: accuracy of the measurement, inclusion of all key dimensions of teachers' professional practice, consistency with the set goals of the exercise, adaptation to the needs of those who will use the results, cost-effectiveness, and practical feasibility. While many countries have reached consensus regarding the areas to be appraised, there is much contention about the instruments for collecting evidence on teachers' practice. Since the process of gathering evidence about a teacher's performance may influence the appraisal results, the choice of instruments is of chief importance in designing and implementing teacher appraisal systems.

While limited resources make trade-offs inevitable, comprehensive teacher appraisal models are more likely to provide a solid basis to appraise teachers. Research points to the importance of combining multiple instruments so as to ensure that all important aspects' of a teacher's performance are given an adequate degree of attention. This is particularly important in high-stakes approaches to teacher appraisal. Different appraisal instruments such as classroom observation, interviews with the teacher, teacher self-appraisal, student performance data and feedback from students and parents can provide different perspectives on the teacher's work and strengthen the validity and reliability of the overall appraisal. Portfolios can be a useful instrument for bringing together different types of evidence including lesson plans, videotapes of lessons, samples of student work and narrative comments.

Provide support for effective classroom observations

Only if teacher appraisal includes classroom observations, it can ensure that individual weaknesses in teaching approaches are picked up and adequately addressed with suitable professional development. Teaching is at the core of a teacher's professional responsibilities, and it can be improved through observation, constructive feedback and learning opportunities. Hence, the observation of teaching by trained evaluators is

essential to a robust system of teacher appraisal. For classroom observations to be useful to teachers it is preferable that they are frequent and take place in a trusting environment.

Each school needs to develop the internal capacity to conduct high-quality classroom observations, which requires targeted investment in training for school leadership teams, as well as teachers. Such training should focus in particular on gaining a deep understanding of what constitutes "good teaching" and how it can be observed using high-quality standards and criteria for appraisal. The advantage of such investment is that it builds on the professionalism of teachers and school leaders and contributes to their development.

Establish safeguards against simplistic use of student results for teacher appraisal

Evidence of student learning outcomes is an important element in evaluating the success of teachers in helping all students achieve. Using student assessment results to appraise teachers is typically intended to focus attention towards student learning as the primary goal of teaching and to incentivise teachers to raise standards. There has been increased interest across countries in value-added models that are designed to control for the previous results of individual students in order to identify the contribution of an individual teacher to student progress. However, despite the attractiveness of such models, there are a range of statistical, methodological and practical challenges in using student standardised assessment results to appraise individual teachers.

Given the difficulties to ensure that value-added measures provide an accurate and comprehensive picture of individual teacher performance, such measures should not be used as the sole measurement of teacher performance and safeguards should be established against a simplistic use of student test results for high-stakes decisions about teachers. But this does not imply that teachers should be exempted from providing evidence of their students' progress, for example through specific evidence of student progress and portfolios.

Ensure that student feedback to teachers is used for formative purposes

Students are not pedagogical experts and student feedback should not be the sole or primary source of information for teacher appraisal. While student feedback can help identify certain problems in teachers' practices, it cannot replace relevant professional feedback, advice and support by teaching experts. Therefore, the use of student surveys is not recommended for high-stakes accountability purposes in teacher appraisal. Student surveys provide more valuable insights for whole-school evaluation and their use for that purpose should be encouraged (see Chapter 6). Also, they may provide highly useful formative feedback to individual teachers.

There are a number of approaches that can potentially increase the usefulness of student surveys for formative purposes: for example, they should focus on teaching practice rather than the teacher as an individual; include the students' own self-and peer-assessment to allow for analysis of classroom interactions; feature questions on teaching approaches that are known to be relevant for student learning; include information on the general framework for teaching such as materials and physical conditions as well; and be analysed by the students and teacher together with a view to improve the classroom environment and teaching and learning processes.

Capacity

Prepare teachers for their role in appraisal processes

Guaranteeing that teachers are provided with support to understand the appraisal procedures and to benefit from appraisal results is also vitally important. Teachers can benefit from training modules so they know what is expected from them to be recognised as good teachers, and to be prepared to make the best use of the feedback received. Initial teacher education programmes should be informed by existing teaching standards to ensure that when teachers enter the profession they already understand what is important for them to know and be able to do. In this way, by the time teachers begin their own practice they are conversant with existing standards and reflect these in their own teaching. Induction and mentoring for new teachers can further ease the transition between initial education and school-level appraisal processes.

While connections to initial teacher education can help the next generation of teachers to be better prepared for their appraisal, it is equally important to ensure that teachers already on the job have opportunities to learn and fully understand appraisal processes and criteria. Substantial activities for professional development on how to best use appraisal processes should be offered to teachers. Also, beyond punctual workshops, teachers could benefit from the creation of networks where professional learning communities of teachers, school leaders and education administrators build a collective understanding of how to evaluate and improve teaching and learning approaches. This could involve peer exchange, discussing complex challenges, sharing and critiquing of practice and fostering a sense of common direction.

Involving teachers as peer evaluators in the appraisal process can also help them learn about key areas of performance with a view to continuously improving their own practice. It is important that peer evaluators are adequately selected and trained for their role to ensure that they have legitimacy in the eyes of teachers being appraised.

Strengthen the capacity of school leaders for teacher appraisal

Effective operation of teacher appraisal and its contribution to school development will depend to a great extent on the strategic human resource management as well as organisational and pedagogical leadership exercised by school leaders. Teacher appraisal will only succeed in raising educational standards if school principals take direct responsibility for improving the quality of teaching and learning in their schools. School leaders are also more likely to provide informal continuing feedback to teachers throughout the year and not only during the formal appraisal process.

More generally, school leaders are essential to promote the continuous improvement of teaching and learning approaches and to insist on the obligation of all staff to engage in regular professional learning. Therefore the recruitment, training, professional development and appraisal of school leaders should be given great importance. Strategies to strengthening school leadership could include the following components: development of a national education programme for school leaders; support for distributed leadership so as to reduce the burden on school principals and foster leadership capacity across the school; enhancing the performance appraisal of school leaders (see Chapter 7); drawing on the expertise of school principals from highly effective schools and supporting peer learning platforms; and supporting the collaboration of school leaders with a critical friend.

Professional development offers targeted at school leaders should include a focus on school quality assurance and improvement, including school self-evaluation and teacher appraisal. This should involve human resource management, including aspects such as structured interactions with teachers, setting of objectives, linking school objectives to personnel development plans, making use of various sources of information on teaching quality, development of appraisal instruments, and management tools to use appraisal results.

Ensure that designated evaluators are qualified for their role

The success of any teacher appraisal system depends greatly on the competencies of designated evaluators. Evaluators should be trained in standards-based methods for assessing evidence of teacher performance. They need to be able to appraise teachers according to the limited evidence they gather, the criteria of good teaching and the corresponding performance levels, where these are defined. Evaluators should also be prepared to provide constructive feedback to teachers for further practice improvement. Where teacher appraisal has consequences for the individual teacher but is essentially school-based, it would be desirable to establish moderation processes to ensure consistency of school-internal approaches to teacher appraisal. Any evaluators involved in such teacher appraisal processes should have received specific training for this function and be accredited by a relevant organisation.

The following range of characteristics and competencies are likely to be helpful for evaluators in strengthening the effectiveness of appraisal processes: *(i)* background in teaching; *(ii)* knowledge of educational evaluation theories and methodologies; *(iii)* knowledge of concepts of teaching quality; *(iv)* familiarity with systems and procedures of educational and school quality assurance, including the role of teaching quality in school quality; *(v)* understanding of instrument development, including reliability and validity of observation and other assessment tools; *(vi)* awareness of the psychological aspects of appraisal; *(vii)* expertise with the quantitative rating of an assessment; and *(viii)* mastering of appraisal-related communication and feedback skills.

Build central expertise to continuously improve teacher appraisal policies and practices

Teacher appraisal is eminently a technical matter and has a lot to benefit from worldwide evidence on best policies and practices. Some countries bring together educational researchers and distinguished teachers into an advisory group to monitor and guide the implementation of teacher appraisal. Such group is in a good position to recognise effective appraisal practices, to keep abreast of relevant research developments and, as a result, to provide advice based on sound evidence. Another option is to establish requirements for national teacher appraisal systems to be adequately informed by scientific evidence.

Use of results

Ensure that teacher appraisal feeds into professional development and school development

Teacher appraisal is unlikely to produce effective results if it is not appropriately linked to professional development. In order for a vibrant programme of professional development to be established and sustained, it must be based on a culture of professional

inquiry. There must be a recognised and explicitly stated norm that recognises the great complexity of good teaching, and insists, therefore, on the professional obligation of every teacher to be engaged in a career-long quest of improved practice. This culture needs to go along with an adequate provision of professional development. It is important that available professional development fits identified needs. For example, if feedback to teachers is provided in relation to the criteria outlined in teaching standards, then professional development activities could be organised around those criteria and be managed locally.

Individual teacher development, in turn, needs to be associated with school development if the improvement of teaching practices is to meet the school's needs. To be most effective, professional development programmes should be coordinated at the school level, so that teachers are aware of the learning goals pursued by their colleagues and potential areas for collaboration. Such joint efforts can contribute to establishing learning communities. Schools that associate the identified individual needs with the school priorities, and that also manage to develop the corresponding professional development activities, are likely to perform well. Schools can learn from the strengths of effective teachers and implement professional development programmes that respond to their weaknesses. Given the important role that school leaders play in linking teacher appraisal results with individual and school development, the adequate preparation of school leaders for this task should be given great importance (see above).

Establish feedback loops between teacher appraisal systems and initial teacher education

Just as individual teachers can improve their teaching when they have identified their relative areas of strength and weakness, so too can initial teacher education programmes improve their approaches when they are informed of the success of their graduates. If initial teacher education and teacher appraisal processes are based on the same teaching standards, and the results of teacher appraisal are fed back to the preparing institutions, teacher education programmes can be strengthened accordingly.

The ideal time for such feedback is during a teacher's initial years, when it is likely that their skills are a direct function of what they would have learned during their preparation. For example, a survey to be completed by all teachers in their first two years would enable teacher education programmes to collect important information about where teachers feel they have been most and least successful and where they wish they would have had more training.

In addition, collecting similar information from school leaders enables teacher education programmes to determine whether the teachers they are preparing are meeting the needs of the schools in the contexts in which they are beginning their teaching careers. This information can then be used by teacher education programmes to make needed curricular adjustments. Such mechanisms also strengthen avenues of communication from schools directly to teacher education programmes.

Establish links between teacher appraisal and career advancement decisions

Teachers need to be acknowledged and have their teaching effectiveness recognised. This can be achieved through linking appraisal results to career advancement. As outlined above, this requires the existence of a performance-based career ladder. The formal diversification of the teaching career can help meet school needs and provide more

opportunities and recognition to teachers, including those who wish to remain focused on classroom teaching.

Within such a career structure, the principle of rewarding teachers for exemplary performance can be applied by associating performance levels to the speed at which a teacher advances in the career (within and across career pathways). Establishing linkages between teacher appraisal and career advancement provides an indirect link between teacher performance levels and pay. This is a desirable option as direct links between teacher performance and pay have produced mixed results, according to the research literature.

Consider the use of non-monetary rewards as a complementary tool to recognise teachers

Reward schemes connecting appraisal results directly to bonus pay should be approached with caution, as the overall impact of such payments is mixed and can be contentious and potentially divisive. Rewarding teachers with non-monetary rewards such as time allowances, sabbatical periods, opportunities for school-based research, support for post-graduate study or professional development could be more appealing incentives. It is important that any performance-based rewards be awarded for reasons which teachers and school leaders perceive as fair and valid.

Some general principles for giving out performance include: *(i)* ensuring that all teachers, regardless of educational level and subject, are eligible for performance awards; *(ii)* using multiple measures of teachers' performance to assess their effectiveness, not only in the classroom but as members of a learning community within the school; *(iii)* rewarding teachers for taking on extra work within the school, such as coaching or mentoring new teachers; and *(iv)* acknowledging teacher professional growth through their participation in coursework and extended professional training in their content area. The more objective the process for determining the merit awards, the more accepted it will be among teachers and staff, since it will be seen as a valid recognition of excellent performance.

Ensure that underperformance is identified and adequately addressed

One of the key functions of teacher appraisal is to ensure that all classrooms are in the hands of capable and motivated teachers. This implies that there should be mechanisms to detect weaknesses in teaching practices and to ensure that underperformance is adequately addressed. Responses to underperformance may include opportunities for professional learning, intense assistance and coaching, further appraisal and feedback and, as a potential second step, the possibility to move on teachers that do not respond to professional learning opportunities.

Notes

1. In Finland, the basis for teacher appraisal is defined in the contract between each local government employer (mostly municipalities) and the teacher trade union.
2. Formerly known as the *National Professional Standards for Teachers* developed by the Australian Institute for Teaching and School Leadership (AITSL) as a foundation reform of the National Partnership on Teacher Quality. Further details can be found at *www.teacherstandards.aitsl.edu.au*.
3. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.
4. In Chile, the appraisal occurs every fourth year if the results are satisfactory, otherwise it occurs more often (every year or every second year).

References

Abell, S.K., D.R. Dillon, C.J. Hopkins, W.D. McInerney and D.G. O'Brien (1995), "'Somebody to count on': Mentor/intern relationships in a beginning teacher internship program", *Teaching and Teacher Education,* Vol. 11, No. 2, pp. 173-188.

AFT and NEA (American Federation of Teachers and National Education Association) (2008), *A Guide to Understanding National Board Certification*, Washington, DC.

Anderson, L. and L. Pellicer (2001), *Teacher Peer Assistance and Review*, Corwin Press.

Attinello, J.R., D. Lare and F. Waters (2006), "The value of teacher portfolios for evaluation and professional growth", *NAASP Bulletin,* Vol. 90, No. 2, pp. 132-152.

Avalos, B. and J. Assael (2006), "Moving from resistance to agreement: The case of the Chilean teacher performance evaluation", *International Journal of Educational Research*, Vol. 45, No. 4-5, pp. 254-266.

Beck, R., N. Livne and S. Bear (2005), "Teachers' self-assessment of the effects of formative and summative electronic portfolios on professional development", *European Journal of Teacher Education*, Vol. 28, No. 3, pp. 221-244.

Berry, B., E. Fuller, C. Reeves and E. Laird (2007), "Linking teacher and student data to improve teacher and teaching quality", *Data Quality Campaign,* Washington, DC, pp. 1-16.

Bivona, L. (2012), *Options for Including Teachers of Nontested Grades and Subjects in Performance-Based Compensation Systems*, American Institutes for Research (AIR), Washington, DC.

Bolino, M. and W. Turnley (2003), "Counternormative impression management, likeability, and performance ratings: The use of intimidation in an organizational setting", *Journal of Organizational Behavior*, Vol. 24, No. 2, pp. 237-250.

Bond, L., T. Smith, W. Baker and J. Hattie (2000), *The Certification System of the National Board for Professional Teaching Standards: A Construct and Consequential Validity Study*, NBPTS.

Borman, G. and S. Kimball (2005), "Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps?", *The Elementary School Journal*, Vol. 106, No. 1, pp. 3-20.

Braun, H. (2005a), *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*, Educational Testing Service (ETS).

Braun, H. (2005b), "Value-added modeling: What does due diligence require?", in R.W. Lissitz (ed.), *Value Added Models in Education: Theory and Applications*, JAM Press, Maple Grove, pp. 19-39.

CAESL (Center for Assessment and Evaluation of Student Learning) (2004), "Using student tests to measure teacher quality", *CAESL Assessment Brief*, No. 9.

Campbell, D., B. Melenzyer, D. Nettles and R. Wyman (2000), *Portfolio and Performance Assessment in Teacher Education*, Allyn and Bacon, Needham Heights, Massachusetts.

Cavalluzzo, L. (2004), *Is National Board Certification An Effective Signal of Teacher Quality?*, The CNA Corporation, Alexandria, Virginia.

Chamberlin, R., T. Wragg, G. Haynes and C. Wragg (2002), "Performance-related pay and the teaching profession: A review of the literature", *Research Papers in Education*, 17(1), pp. 31-49.

Chilean Ministry of Education (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Chile*, *www.oecd.org/edu/evaluationpolicy*.

Coggshall, J.G., E. Behrstock-Sherratt and K. Drill (2011), *Workplaces that Support High-Performing Teaching and Learning: Insights from Generation Y Teachers*, American Federation of Teachers and American Institutes for Research, Washington, DC.

Cohen, C. and J. Rice (2005), *National Board Certification as Professional Development: Design and Cost*, National Board for Professional Teaching Standards (NBPTS).

Daley, G. and L. Kim (2010), *A Teacher Evaluation System that Works,* Working Paper, National Institute for Excellence in Education, August 2010.

Danielson, C. (2011), "Evaluations that help teachers learn", *Educational Leadership*, December 2010/ January 2011, pp. 35-39.

Danielson, C. (2001), "New trends in teacher evaluation", *Educational Leadership*, Vol. 58, No. 5, pp. 12-15.

Danielson, C. (1996, 2007), *Enhancing Professional Practice: A Framework for Teaching*, 1st and 2nd Editions, Association for Supervision and Curriculum Development (ASCD), Alexandria, Virginia.

Danielson, C. and T. McGreal (2000), *Teacher Evaluation to Enhance Professional Practice*, Association for Supervision and Curriculum Development (ASCD), Alexandria, Virginia.

Darling-Hammond, L. (2012), "The right start: Creating a strong foundation for the teaching career", *Phi Delta Kappan*, Vol. 94, No. 3, pp. 8-13.

Darling-Hammond, L. (2001), "Portfolio as practice: The narratives of emerging teachers", *Teaching and Teacher Education*, Vol. 17, pp. 107-121.

Darling-Hammond, L., A. Amrein-Beardsley, E. Haertel and J. Rothstein (2012), "Evaluating teacher evaluation", *Phi Delta Kappan*, Vol. 93, No. 6, pp. 8-15.

Department of Education, Northern Ireland (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Northern Ireland, United Kingdom*, *www.oecd.org/edu/evaluationpolicy*.

Freund, M., V. Kane Russell and C. Kavulic (2005), *A Study of the Role of Mentoring in Achieving Certification by the National Board for Professional Teaching Standards*, National Board for Professional Teaching Standards (NBPTS).

Glazerman, S., D. Goldhaber, S. Loeb, S. Raudenbush, D.O. Staiger and G.J. Whitehurst (2011), *Passing Muster: Evaluating Evaluation Systems*, Brown Center on Education Policy at Brookings, Washington, DC.

Goe, L. (2007), *The Link between Teacher Quality and Student Outcomes: A Research Synthesis*, National Comprehensive Center for Teacher Quality, 2007.

Goe, L., C. Bell and O. Little (2008), *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis,* National Comprehensive Center for Teacher Quality, Washington, DC.

Goe, L., K. Biggers and A. Croft (2012), *Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning*, National Comprehensive Center for Teacher Quality, Washington, DC.

Goldhaber, D. and E. Anthony (2007), "Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching", *The Review of Economics and Statistics*, Vol. 89, No. 1, pp. 134-150.

Goldstein, J. (2007), "Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review, *American Journal of Education*, Vol. 113, pp. 479-508.

Goldstein, J. and P.A. Noguera (2006), "A thoughtful approach to teacher evaluation", *Educational Leadership,* Vol. 63(6), pp. 31-37.

Hanover Research (2011), *A Survey of Race to the Top Teacher Evaluation Systems*, Washington, DC.

Hanushek, E. (1986) "The economics of schooling: Production and efficiency in public schools", *Journal of Economic Literature*, Vol. 24, No. 3, pp. 1141-1177.

Harris, W.J. and S.V. MacKenzie (2007), *National Board Teacher Certification in Maine: An Exploratory Study*, Maine Education Policy Research Institute, Orono, Maine.

Harris, D. and T. Sass (2007), "The effects of NBPTS-certified teachers on student achievement", *Working Paper No. 4*, Center for Analysis of Longitudinal Data in Education Research (CALDER).

Harvey-Beavis, O. (2003), "Performance-based rewards for teachers: A literature review", paper distributed at the third workshop of participating countries in the OECD activity "Attracting, Developing and Retaining Effective Teachers", Athens, 4-5 June, *www.oecd.org/edu/teacherpolicy*.

Heilbronn, R., C. Jones, S. Bubb and M. Totterdell (2002), "School-based induction tutors: A challenging role", *School Leadership & Management,* Vol. 22, No. 4, pp. 371-87.

Heneman, H., A. Milanowski and S. Kimball (2007), *Teacher Performance Pay: Synthesis of Plans, Research, and Guidelines for Practice*, Consortium for Policy Research in Education (CPRE) Policy Briefs RB-46.

Heneman, H., A. Milanowski, S. Kimball and A. Odden (2006), "Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay", *Consortium for Policy Research in Education (CPRE) Policy Briefs,* RB-45.

Hess, F. and M. West (2006), *A Better Bargain: Overhauling Teacher Collective Bargaining for the 21st Century*, Program on Education Policy and Governance, Harvard University, Cambridge, Massachusetts.

Hill, H. and C. Herlihy (2011), "Prioritizing teaching quality in a new system of teacher evaluation", *Education Outlook*, No. 9, American Enterprise Institute for Public Policy Research.

Hobson, A.J. (2009), "On being bottom of the pecking order: Beginner teachers' perceptions and experiences of support", *Teacher Development: An International Journal of Teachers' Professional Development,* Vol. 13, No. 4, pp. 299.

Hobson, A.J., P. Ashby, A. Malderez and P.D. Tomlinson (2009), "Mentoring beginning teachers: What we know and what we don't", *Teaching and Teacher Education: An International Journal of Research and Studies*, Vol. 25, No. 1, pp. 207-216.

Ingvarson, L. and J. Hattie (2008), *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards*, Elsevier/JAI, Amsterdam/Boston.

Ingvarson, L., E. Kleinhenz and J. Wilkinson (2007), *Research on Performance Pay for Teachers*, Australian Council for Educational Research (ACER).

Isoré, M. (2009), "Teacher evaluation: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 23, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers.*

Jackson, C.K. and E. Bruegmann (2009), *Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers*, National Bureau of Economic Research, Cambridge, Massachusetts.

Jacob, B. and L. Lefgren (2008), "Can principals identify effective teachers? Evidence on subjective performance evaluation in education", *Journal of Labor Economics*, Vol. 26, No. 1, pp. 101-136.

Jacob, B. and L. Lefgren (2005), "What do parents value in education: An empirical investigation of parents' revealed preferences for teachers", *NBER Working Paper*, No. 11494.

Jensen, B. and J. Reichl (2011), *Better Teacher Appraisal and Feedback: Improving Performance*, Grattan Institute, Melbourne.

Kane, T.J. and D.O. Staiger (2012), *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*, MET Project Policy & Practice Brief, Bill and Melinda Gates Foundation, Seattle, Washington.

Kane, T. and D. Staiger (2002), "Volatility in school test scores: Implications for test-based accountability systems", *Brooking Papers on Education Policy*, Washington, DC.

Kane, T.J., E.S. Taylor, J.H. Tyler and A.L. Wooten (2010), *Identifying Effective Classroom Practices Using Student Achievement Data*, National Bureau of Economic Research, Cambridge, Massachusetts.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and Jungha Park (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

Klecker, B. (2000), "Content validity of preservice teacher portfolios in a standards-based program", *Journal of Instructional Psychology*, Vol. 27, No. 1, pp. 35-38.

Klinger, D.A., L.M. Shulha and C. DeLuca (2008), "Teacher evaluation, accountability, and professional learning: The Canadian perspective", *Rev. Pensamiento Educativo*, Vol. 43, pp. 209-222.

Larsen, M., with C. Lock and M. Lee (2005), *Professional Certification and Entry-to-Practice Assessments: A Report for the Teaching Policy and Standards Branch*, Ontario Ministry of Education, Ontario, Canada.

Lee, C.K. and M.Y. Tan (2010), "Rating Teachers and Rewarding Teacher Performance: The Context of Singapore", Paper presented at APEC Conference on Replicating Exemplary Practices in Mathematics Education, Koh Samui, Thailand, 7-12 March 2010.

Lustick, D. and G. Sykes (2006), "National Board Certification as professional development: What are teachers learning?", *Education Policy Analysis Archives*, Vol. 14, No.5.

Mansvelder-Longayroux, D., D. Beijaard and N. Verloop (2007), "The portfolio as a tool for stimulating reflection by student teachers", *Teaching and Teacher Education*, Vol. 23, No. 1, pp. 47-62.

Margo, J., M. Benton, K. Withers and S. Sodha (2008), *Those Who Can?*, Institute for Public Policy Research (IPPR) Publications.

Marshall, K. (2012), "Let's cancel the dog-and-pony show", *Phi Delta Kappan*, Vol. 94, No. 3, pp. 19-23.

McCaffrey, D.F., J.R. Lockwood, D. Koretz and L. Hamilton (2003), *Evaluating Value-Added Models for Teacher Accountability*, Rand Corporation.

McColskey, W. and J. Stronge (2005), *A Comparison of National Board Certified Teachers and non-National Board Certified Teachers: Is there a Difference in Teacher Effectiveness and Student Achievement?*, National Board for Professional Teaching Standards (NBPTS).

Mead, S., A. Rotherham and R. Brown (2012), *The Hangover: Thinking About the Unintended Consequences of the Nation's Teacher Evaluation Binge*, Teacher Quality 2.0, American Enterprise Institute, Special Report 2, September.

Meyer, R. and M. Christian (2008), *Value-Added and Other Methods for Measuring School Performance,* NCPI Conference 2008, National Center on Performance Incentives.

Milanowski, A. (2005), "Split roles in performance evaluation: A field study involving new teachers" *Journal of Personnel Evaluation in Education,* Vol. 18, No. 3, pp. 153-169.

Milanowski, A. (2004) "The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati", *Peabody Journal of Education*, Vol. 79, No. 4, pp. 33-53.

Milanowski, A. and S. Kimball (2003), "The framework-based teacher performance assessment systems in Cincinnati and Washoe", *CPRE Working Paper Series*, TC-03-07.

Mohrman, A., S. Mohrman and A. Odden (1996), "Aligning teacher compensation with systemic school reform: Skill-based pay and group-based performance awards", *Educational Evaluation and Policy Analysis*, 18(1), pp. 51-71.

Morris, A. (2011), "Student standardised testing: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 65, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

NBPTS (National Board for Professional Teaching Standards) (2007), *A Research Guide on National Board Certification of Teachers*, National Board for Professional Teaching Standards, Arlington, Virginia, 2007.

National Research Council (2010), *Preparing Teachers: Building Evidence for Sound Policy*, Committee on the Study of Teacher Preparation Programs in the United States, T.N.A. Press (Ed.), Division of Behavioral and Social Sciences and Education, Washington, DC.

Norwegian Directorate for Education and Training (2008), *Head Teacher Competence – Expectations and Demands Core Document and Point of Departure for Announcement of Tender (Attachment 2)*, translated from Norwegian.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

O'Day, J. (2002), "Complexity, accountability, and school improvement", *Harvard Educational Review,* Vol. 72, No. 3, pp. 293-329.

OECD (2012), *The Experience of New Teachers: Results from TALIS 2008*, OECD Publishing, Paris.

OECD (2010), *Improving Schools: Strategies for Action in Mexico*, OECD Publishing, Paris.

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS,* OECD Publishing, Paris.

OECD (2006), "The teaching workforce: Meeting aspirations and enhancing motivation", in *Education Policy Analysis 2005*, OECD Publishing, Paris.

OECD (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OECD Publishing, Paris.

Ofsted (Office for Standards in Education) (2006), *The Logical Chain: Continuing Professional Development in Effective Schools*, OFSTED Publications No. 2639, United Kingdom.

Papay, J.P. (2012), "Refocusing the debate: Assessing the purposes and tools of teacher evaluation", *Harvard Educational Review*, Vol. 82, No. 1.

Pecheone, R.L., M.J. Pigg, R.R. Chung and R.J. Souviney (2005), "Performance assessment and electronic portfolios: Their effect on teacher learning and education", *The Clearing House,* Vol. 78, No. 4, pp. 164-176.

Peterson, K. (2000), *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*, 2nd Edition, Corwin Press, Thousand Oaks, California.

Peterson, K.D. (1987), "Teacher evaluation with multiple and variable lines of evidence," *American Educational Research Journal,* Vol. 24, No. 2, pp. 311-317.

Peterson, K., C. Wahlquist and K. Bone (2000), "Student surveys for teacher evaluation", *Journal of Personnel Evaluation in Education*, Vol. 14, No. 2, pp. 135-153.

Peterson, K., C. Wahlquist, J. Esparza Brown and S. Mukhopadhyay (2003), "Parents surveys for teacher evaluation", *Journal of Personnel Evaluation in Education*, Vol. 17, No. 4, pp. 317-330.

Petty, T. (2002), *Identifying the Wants and Needs of North Carolina High School Mathematics Teachers for Job Success and Satisfaction*, National Board for Professional Teaching Standards (NBPTS).

Pont, B., D. Nusche and H. Moorman (2008), *Improving School Leadership – Volume 1: Policy and Practice*, OECD Publishing.

Ramirez, A. (2001), "How merit pay undermines education", *Educational Leadership*, Vol. 58(5), pp. 16-20.

Randi, J. and K.M. Zeichner (2004), "New visions of teacher professional development", in M.A. Smylie and D. Miretzky (eds.), *Developing the Teacher Workforce: 103rd Yearbook of the National Society for the Study of Education (Part I)*, The University of Chicago Press, Chicago.

Richardson, R. (1999), *Performance Related Pay in Schools: An Assessment of the Green Papers*, report prepared for the National Union of Teachers, The London School of Economics and Political Science.

Robinson, V. (2007), "School leadership and student outcomes: Identifying what works and why", *ACEL Monograph Series*, No. 41, Australian Council for Educational Leaders.

Rockoff, J. and C. Speroni (2011), "Subjective and objective evaluations of teacher effectiveness: Evidence from New York City, *Labour Economics,* Vol. 18, pp. 687-696.

Rosenkvist, M. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, *www.oecd.org/edu/workingpapers*.

Sanders, W., J. Ashton and P. Wright (2005), *Comparison of the Effects of NBPTS Certified Teachers with Other Teachers on the Rate of Student Academic Progress*, National Board for Professional Teaching Standards (NBPTS).

Santiago, P. and F. Benavides (2009), *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*, OECD, Paris, *www.oecd.org/edu/evaluationpolicy.*

Santiago, P., F. Benavides, C. Danielson, L. Goe and D. Nusche (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile*, OECD Publishing, *www.oecd.org/edu/evaluationpolicy.*

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, *www.oecd.org/edu/evaluationpolicy.*

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012b), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, *www.oecd.org/edu/evaluationpolicy*.

Sartain, L., S.R. Stoelinga and E. Krone (2010), *Rethinking Teacher Evaluation: Findings from the First Year of the Excellence in Teacher Project in Chicago Public Schools,* Consortium on Chicago Public Schools Research at the University of Chicago, Chicago, Illinois.

Schochet, P.Z. and H.S. Chiang (2010), *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark*, OECD Publishing, *www.oecd.org/edu/evaluationpolicy*.

Smith, T., B. Gordon, S. Colby and J. Wang (2005), *An Examination of the Relationship between Depth of Student Learning and National Board Certification Status*, Office for Research on Teaching, Appalachian State University.

Steele, J.L., L.S. Hamilton and B.M. Stecher (2010), *Incorporating Student Performance Measures into Teacher Evaluation Systems,* RAND, Santa Monica, California.

Storey, A. (2000), "A leap of faith? Performance pay for teachers", *Journal of Education Policy*, 15(5), pp. 509-23.

Stronge, J. and P. Tucker (2003), *Handbook on Teacher Evaluation: Assessing and Improving Performance*, Eye On Education Publications.

Sykes, G. and S. Winchell (2010), "Assessing for teacher tenure", in M. Kennedy (ed.), *Teacher Assessment and the Quest for Teacher Quality*, Jossey-Bass, San Francisco, pp. 201-224.

Taylor, E.S. and J.H. Tyler (2011), "The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers", *NBER Working Paper Series,* No. 16877, National Bureau of Economic Research, Cambridge, Massachusetts.

TDA (Training and Development Agency for Schools) (2007), *Professional Standards for Teachers*, TDA, London, *www.tda.gov.uk/upload/resources/pdf/s/standards_a4.pdf*.

Timperley, H. (2011), *Realising the Power of Professional Learning*, Open University Press, London.

Tucker, P., J. Stronge and C. Gareis (2002), *Handbook on Teacher Portfolios for Evaluation and Professional Development*, Eye on Education, Larchmont, New York.

Tucker, P.D., J.H. Stronge, C.R. Gareis and C.S. Beers (2003), "The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference?", *Educational Administration Quarterly,* Vol. 39, No. 5, pp. 572-602.

Vandervoort, L., A. Amrein-Beardsley and D. Berliner (2004), "National Board certified teachers and their students' achievement", *Education Policy Analysis Archives*, Vol. 12, No. 46.

van der Schaaf, M.F., K.M. Stokking and N. Verloop (2008), "Teacher beliefs and teacher behaviour in portfolio assessment, *Teaching & Teacher Education,* Vol. 24, No. 7, pp. 1691-1704.

Wei, R.C., L. Darling-Hammond, A. Andree, N. Richardson and S. Orphanos (2009), *Professional Learning in the Kearning Profession: A Status Report on Teacher Development in the U.S. and Abroad*, National Staff Development Council, Dallas, Texas.

Weinstein, T. L. and K. S. Struthers (2012), "Similar demands, different responses: Teacher evaluation in the United Kingdom and Singapore", *Educational Policy Analysis and Strategic Research*, Vol. 7, No. 1.

Williams, A. and S. Prestage (2002), "The induction tutor: Mentor, manager or both?", *Mentoring & Tutoring: Partnership in Learning,* Vol. 10, No. 1, pp. 35.

Annex 5.A Features of teacher appraisal frameworks

The tables below provide information on features of teacher appraisal frameworks in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. The information was supplied by countries through a questionnaire specifically developed by the OECD Review.

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

General notes

Australia: Australia is a federation of eight states and territories. There are differences in policy frameworks for evaluation and assessment across states and territories as well as between public (government) and private (non-government) schools. From 2013, national standards for teachers are being introduced across all states and territories providing a national measure of teaching practice.

Belgium (Fl., Fr.): In Belgium, education policy is the responsibility of each Community. The terms "national" and "central", therefore, refer to the highest education authorities (Ministries of Education) of the Flemish and French Communities of Belgium.

Belgium (Fl.): For public schools, the school organising bodies are typically the central education authority (Flemish Community) and provincial/regional and local education authorities (provinces, cities, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Belgium (Fr.): For public schools, the school organising bodies (education networks) are typically the central education authority (French Community) and provincial and local education authorities (provinces, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Canada: Canada comprises ten provinces and three territories. Provincial/territorial education authorities refer to the highest level of educational authorities in Canada, as there is no federal/central department of education. There are differences in policy frameworks for evaluation and assessment across provinces and territories.

Chile: For public schools, the school organising bodies (sustainers) are typically local education authorities (municipalities).

Czech Republic: For public schools, the school organising bodies are typically local education authorities at ISCED levels 1 and 2 and regional education authorities at ISCED level 3.

Hungary: For public schools, school organising bodies (maintainers) are typically local and regional education authorities. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

Ireland: School boards of management comprise members external to the school such as representatives of the patron and of the local community as well as members internal to the school such as the principal, teacher representatives and parent representatives.

Netherlands: In principle, all schools are government funded. Depending on their denomination, they can have a private (religious or pedagogy-based) or public character. For public schools, school organising bodies (competent authorities) can be local educational authorities (municipal authorities), a local governing committee with transferred powers of the municipality, or a public foundation or corporation. School organising bodies for private schools can be groups of parents, foundations or corporations.

New Zealand: School Boards of Trustees typically comprise elected members from the school community, the principal, a staff representative and a student representative (in secondary schools).

Spain: Responsibilities for education are shared between the central education authority (Ministry of Education, Culture and Sport [*Ministerio de Educación, Cultura y Deporte*]) and state education authorities (Regional Ministries or Departments of Education of the Autonomous Communities [*Comunidades Autónomas*]). The central education authority executes the general guidelines of the government on education policy and regulates the basic elements or aspects of the system. The Autonomous Communities develop the central regulations and have executive and administrative competences for managing the education system in their own territory. State education authorities refer to education authorities at the highest level of the Autonomous Communities. Throughout the tables, the Autonomous Communities are referred to as "state education authorities".

General notes for teacher appraisal

Australia: Australia is a federation of eight states and territories. There are differences in teacher appraisal systems between states and territories and also between public (government) and private (non-government) schools. For most private schools, the teacher appraisal practices are set at the school level. From 2013 the implementation of the Australian Teacher Performance and Development Framework will ensure that every teacher receives regular, appropriate and constructive feedback on their performance.

Canada: Canada is a federation of ten provinces and three territories. There are differences in teacher appraisal across provinces and territories.

Denmark: There is no central policy framework for the mandatory appraisal of teachers at ISCED levels 1 and 2. However, appraisal processes can be carried out at a local level, as illustrated by the OECD Teaching and Learning International Survey (TALIS). In general, teachers receive appraisal or feedback from their school leader once a year.

Norway: While there is no central policy framework for the appraisal of teachers in Norway, teacher appraisal can be organised at a local level.

Sweden: Besides the central policy frameworks for teacher appraisal described in Tables 5.A.2 and 5.A.3, teachers are also appraised at the discretion of their school principal according to regulations set by local education authorities.

Table 5.A.1 Teacher appraisal for performance management (2012)

This table describes the appraisal of individual in-service teachers to make a judgement about their work and performance using objective criteria for performance management purposes, i.e. the formal regular appraisal process designed to ensure that individual and organisational goals are met. As such, performance management is part of wider processes and systems for measuring, monitoring and enhancing the performance of teachers.

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Australia	Regular appraisal	All teachers	Mandatory periodic (frequency varies nationally, but generally annually); In case of performance problems	State education authorities or governments	School principal; member of school leadership other than school principal; supervisor; peer evaluator at the same level	State teaching standards; a description of the general and professional duties of teachers; code of conduct From 2013: National teaching standards; national performance and development framework	Planning and preparation; instruction; classroom environment; professional development; links to the community	Classroom observation; dialogue with the teacher; teacher self-appraisal; teacher portfolio	No	Yes, it systematically results in a professional development plan	Yes, appraisal results influence the speed at which a teacher progresses in the career structure and salary scale	Yes (to the extent that it allows the teacher to progress on the salary scale)	None	Further appraisal; compulsory training; salary increment withheld; permanent contract not granted; dismissal; suspension; transfer
	Registration	All teachers	Mandatory non-periodic; mandatory periodic (some states/territories require renewal after a specific time)	Teacher professional organisation; state education authorities or governments	Teacher professional organisation	National teaching standards; state teaching standards; code of conduct	Competencies as described in the National Professional Standards for Teachers (includes planning and preparation; instruction; classroom environment; professional development; conduct)	Teacher portfolio; dialogue with the teacher	Yes (registration status)	Yes, it systematically results in a professional development plan	Yes, appraisal determines eligibility to seek employment as a teacher	No	None	Further appraisal; compulsory training; registration suspended or withdrawn

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Austria	Regular appraisal	All teachers (except teachers on probation)	Mandatory non-periodic; at the discretion of school principal or school board[1]	Central education authority	School principal. Second evaluator in the case of a dispute/appeal in the appraisal process or a complaint about a teacher: state or regional education authority	None	At the discretion of evaluators (may include planning and preparation; instruction; classroom environment)	At the discretion of evaluators (may include classroom observation; dialogue with the teacher)	No	Yes, it may result in a professional development plan for some teachers (teachers underperforming)	No	No	None	Transfer (rarely applied) Suspension/ dismissal (only in very severe cases)
	Contract renewal	Teachers on fixed-term contracts in public and government-dependent private schools only	Mandatory periodic (annually)	Central education authority	School principal. Second evaluator in the case of a dispute/appeal in the appraisal process or a complaint about a teacher: state or regional education authority	None	At the discretion of evaluators (may include planning and preparation; instruction; classroom environment)	At the discretion of evaluators (may include classroom observation; dialogue with the teacher)	Yes (2 levels: pass; fail)	No	No	No	None	Contract renewal/ permanent contract not granted (rarely applied)
	Contract renewal/ completion of probation for teachers on fixed-term contracts	Teachers on fixed-term contracts (ISCED level 2 [academic programmes] and 3)	Mandatory (ongoing) during probationary period (1 year)	Central education authority; state education authorities	School principal; Inspectorate	None	Planning and preparation; instruction; classroom environment; professional development; contribution to school development	Classroom observation; dialogue with evaluators; teacher portfolio	Yes (3 levels)	No	Yes, appraisal results influence decisions on a) access to contract; b) renewal of a fixed-term contract; or c) access to a permanent position	No	None	Failure to pass probationary period (rarely applied)
Belgium (Fl.)	Regular appraisal	All teachers	Mandatory periodic (at least every 4 years)	Central government	First evaluator: school principal Second evaluator: school organising bodies	National teaching standards	Planning and preparation; instruction; classroom environment; professional development; links to the community	Classroom observation; teacher self-appraisal; teacher portfolio; dialogue with the teacher	Yes (2 levels: pass; fail)	Yes, it is expected to influence professional development activities	Yes, the job description can be adapted based on appraisal results	No	None, but school principals may at times reward teachers with a change in work responsibilities	Further appraisal; compulsory training

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Belgium (Fr.)	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Canada	Regular appraisal once every 5 years or in case of performance concerns	All teachers	Mandatory periodic; experienced teachers formally appraised once every 5 years; at the discretion of school principal in case of performance concerns	Provincial/territorial education authorities or governments	School principal	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development; contribution to school development	Classroom observation; dialogue with the teacher; peer collaboration; parent survey	Varies across provinces/ territories	Yes, it is expected to influence professional development activities	No	No	None	At the discretion of the evaluator; withdrawal or inaccessibility to the priority list; termination of employment
	Appraisal for professional development	All teachers	Mandatory periodic	Provincial/territorial education authorities or governments	School principal	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development; contribution to school development	Classroom observation; dialogue with the teacher; peer collaboration; parent survey	Varies across provinces/ territories	Yes, it is expected to influence professional development activities	Yes, appraisal results influence the speed at which a teacher progresses in the career structure; appraisal results will/will not give access to a priority list, intended to grant contracts	No	None	Withdrawal; inaccessibility to the priority list
	Appraisal of teachers during probation (for formative and summative purposes)	Teachers on probation	In relation to decision on employment status	Provincial/territorial education authorities or governments	School principal; superintendent	Provincial teacher standards or competencies	Instruction (engages students; applies creativity and innovation); classroom environment (inclusive learning environment); professional responsibilities	Classroom observation; dialogue with the teacher; peer collaboration; parent survey	Varies across provinces/territories (e.g. 2 levels: pass, fail; performing in a satisfactory manner, performing in a non-satisfactory manner)	Yes, it is expected to influence professional development activities	Yes	No	Recognition; professional learning; employment opportunities	Compulsory training; contract not renewed; permanent contract not granted; loss of certification

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Chile	Regular appraisal	Teachers active in the classroom in public schools only	Mandatory periodic (4 years if results are satisfactory, 1 or 2 years if results are unsatisfactory)	Central education authority or government	External accredited evaluators; school principal; teacher from another school; Local Assessment Committee (composed of peer evaluators of the same local education authority [municipality])[1]	National teaching standards	Planning and preparation;instruction; classroom environment;competencies for student assessment	Teacher self-appraisal; third party reference report by the school leader; dialogue with the peer evaluator; teacher portfolio; classroom observation (class recording)	Yes (4 levels)	Yes, it systematically results in a professional development plan for teachers obtaining a "basic" or "poor" performance rating	No	Yes, teachers with a satisfactory appraisal result may opt for an additional appraisal. In case of a second satisfactory performance, teachers receive salary increase between 5 and 25% of the base salary on a fixed-term basis (between 2 and 4 years).	None	Compulsory training; dismissal
Czech Republic	Regular appraisal	All teachers[2]	In relation to decision on employment status; at the discretion of the school principal	Central education authority or government; school principal	School principal	School internal regulations	Varies across schools depending on school internal regulations and may include: planning and preparation; instruction; classroom environment; professional development; links to the community	Varies across schools depending on school internal regulations and may include classroom observation; dialogue with the teacher; teacher portfolio	No	Varies across schools depending on school internal regulations and may include a professional development plan	No	Varies across schools depending on school internal regulations and may include a pay allowance is provided for good performance (e.g. annually for activities with the class)	Varies across schools depending on school internal regulations and may include support for post-graduate study, extra opportunities for professional development, public recognition	Varies across schools depending on school internal regulations and may include salary increment withheld; further appraisal; compulsory training; permanent contract not granted

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Czech Republic (*continued*)	Appraisal for promotion	All teachers[2]	In relation to decision on employment status; at the discretion of the school principal	Central education authority or government; school principal	School principal	School internal regulations	Professional development; contribution to school development; additional tasks and responsibilities depending on a teacher's professional status as defined through the Act on Pedagogical Employees (e.g. pedagogical advisor)	Teacher portfolio	No	Yes, it systematically results in a professional development plan, if connected with a certain professional status (e.g. pedagogical advisor)	Yes, appraisal results influence the speed at which a teacher progresses in the career structure	Yes, a pay allowance is provided for good performance (e.g. annually depending on activities with the class)	Support for post-graduate study; extra opportunities for professional development (non-mandatory)	Salary increment withheld; permanent contract not granted
Denmark	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Estonia	Appraisal for promotion	All teachers	Voluntary	Central education authority	Special commission	A description of special tasks and roles	Professional development; work efficiency; fulfilment of qualification requirements	Evaluation of documents provided as part of teacher self-appraisal; judgement of the teacher's work by school leaders	No	No	Yes	Yes	Public recognition	Further appraisal
Finland	None	a	a	a	a	a	a	a	a	a	a	a	a	a
France	Regular appraisal	Public schools: permanent teachers Private schools: teachers as salaried employees	Mandatory periodic;[3] voluntary (at the initiative of teachers themselves); in case of underperformance	Central education authority (for appraisal procedures and rubrics at ISCED levels 2 and 3); central government (per decree for appraisal consequences and career advancement); individual evaluators are in charge of choosing certain instruments	General inspectorate; school principal	National norms and standards (through decrees and circulars); school development plan or school project	Instruction; classroom environment; contribution to school development (teacher commitment); work ethic (presence at school; punctuality; authority; interpersonal skills); compliance with curriculum and reforms	Classroom observation; dialogue with the teacher; teacher self-appraisal (preparatory questionnaire); evaluation of documents provided by the teacher and students	Yes (ISCED level 1: range of scores and descriptive ratings depending on the authority [e.g. poor; fair], ISCED levels 2 and 3: range of scores)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence the speed at which a teacher progresses in the career structure and salary scale	Yes (to the extent that it allows the teacher to progress on the career structure and salary scale)	Horizontal or vertical promotion	Further appraisal; permanent contract not granted; deferral of promotion

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Hungary	Regular appraisal	All teachers	Mandatory periodic (about every 2 years)	Central education authority or government; schools[4]	School principal; evaluators specified in the Quality Assurance Programme of the school	School internal regulations based on legal requirements	Varies across schools depending on school internal regulations, but typically a wide variety of aspects	Teacher self-appraisal; judgement by the school principal	Yes (4 levels: exceptionally suitable, suitable, less suitable, not suitable)	Varies across schools depending on school internal regulations, but typically yes	Yes, appraisal results influence the speed at which a teacher progresses in the career structure and salary scale[5]	Yes, appraisal results may bring on a supplement to the base salary with a performance bonus[6]	Varies across schools depending on school internal regulations	Varies across schools depending on school internal regulations and decision by school principal as employer-related responsibility
Iceland	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Ireland	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Israel	Regular appraisal	All teachers[7]	Mandatory periodic (every 3 years)	Central education authority	School principal	National teaching standards	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to the community	Classroom observation; dialogue with the teacher; teacher self-appraisal; teacher portfolio	Yes (5 levels)	Yes, it is expected to influence professional development activities	No	No	None (under discussion)	Further appraisal; deferral of promotion
	Appraisal for promotion	Permanent teachers in public and government-dependent private schools only (senior level only)[7]	In relation to decision on employment status	Central education authority	School principal	National teaching standards	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to the community	Classroom observation; dialogue with the teacher; teacher self-appraisal; teacher portfolio	Yes (5 levels)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decisions about promotion	No	None (under discussion)	Further appraisal; deferral of promotion; salary increment withheld
Italy	None	a	a	a	a	a	a	A	a	a	a	a	a	a

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Korea	Regular appraisal for professional development	All teachers (for professional development)	Mandatory periodic (annually)	Central education authority	Peer evaluators at the same school	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; student guidance	Classroom observation; student surveys; parent surveys	Yes (5 levels)	Yes, it systematically results in a professional development plan	No	No	Sabbatical periods; extra opportunities for professional development	Compulsory training
	Regular appraisal for promotion	Teachers in public schools only (for performance management/ promotion)	Mandatory periodic (annually)	Central education authority	School principal; peer evaluator at the same school	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development; links to the community; student guidance	Observation of performance	Yes (4 levels)	No	Yes, appraisal results influence the decision about promotion	No	None	None
Luxembourg	Appraisal for transfer to another school	All teachers at ISCED level 1 only[8]	In relation to decision on employment status (teacher's request to change school)	Central education authority	Inspector	A description of the general and professional duties of teachers (e.g. job description)	Planning and preparation; instruction; classroom environment	Dialogue with the teacher; classroom observation	Yes (2 levels: pass; fail)	No	No	No	None	Further appraisal
Mexico	Regular appraisal as part of performance management [SLED][9]	Teachers at ISCED level 3, public central level schools only[10]	Mandatory periodic (annually)	Central education authority or government	Peer evaluator at the same school	A description of the general and professional duties of teachers; school development plan or school project	Planning and preparation; professional development; contribution to school development	Teacher self-appraisal; student surveys; classroom observation by peer evaluator at the same school	No[11]	Yes, it systematically results in a professional development plan	Yes, appraisal results influence decision about access to a permanent position	No	None	Permanent contract not granted
	Universal appraisal system for diagnosis and professional development[12]	All teachers at ISCED levels 1 and 2	Mandatory periodic (annually)	Central education authority	Central education authority	A description of the general and professional duties of teachers; code of conduct	Planning and preparation; professional development; knowledge of subject	Teacher testing; student outcomes (standardised assessment results)	No	Yes, it systematically results in a professional development plan	No	No	None	None

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Mexico (*continued*)	National Exam of Teaching Knowledge and Skills for entry to the profession	All teachers in public schools (on permanent and fixed-term contracts)[13]	Voluntary in relation to decision on employment status (e.g. renewal of contract, conversion or awarding of a permanent contract)	Central education authority; state education authorities or governments; teacher professional organisation (Independent Federalist Evaluation Unit [OEIF])[14]	Central education authority or government	A description of the general and professional duties of teachers; code of conduct	Instruction; planning and preparation; specific intellectual abilities; regulations; management and teaching ethics[15]	Teacher testing	Yes (2 levels: acceptable; not acceptable)	No	Yes, appraisal results influence decision about access to a permanent position	No	None	Permanent contract not granted
	National Continuous Training Exams for In-service Teachers (ENAMS) for diagnosis of teacher competencies[16]	Teachers in public schools only	Voluntary (once per year)	Central education authority	Central education authority; external accredited evaluator (National Assessment Centre for Higher Education [CENEVAL]);[17] school board	None	ISCED level 1: pedagogical knowledge related to the competency-based curriculum and the Comprehensive Reform of Basic Education (RIEB)[18] ISCED levels 1; 2 and 3: subject area knowledge in line with the teacher's educational level; service and position	Teacher testing[19]	No[20]	Yes, it systematically results in a professional development plan	Yes, appraisal results influence decisions about promotion	No	None	None
Netherlands	Regular appraisal	All teachers	Mandatory periodic (ISCED level 1: every 4 years; ISCED levels 2 and 3: every 3 years)[21]	Central employer (National Council of School Boards); school organising bodies (competent authorities)[22]	School principal representing school organising body (competent authority)	National teaching standards	Organisational; pedagogical and subject matter competences; interpersonal competences; teamwork; links to the community; professional development	Extensive descriptions of competencies	Varies across schools depending on school internal regulations	Yes, it is expected to result in a professional development plan	Varies across schools depending on school internal regulations	Varies across schools depending on school internal regulations	Changes in work responsibilities; extra opportunities for in-service professional development	Varies across schools depending on school internal regulations

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
New Zealand	Regular appraisal	All registered teachers	Mandatory periodic (annually)	Central education authority or government	Member of school leadership team; peer evaluator from the same school	National Registration Standards	Planning and preparation; instruction; classroom environment; professional development; links to the community; values; professional leadership; responsiveness to diverse linguistic and cultural backgrounds; analysis and use of assessment information; critical inquiry and problem solving	Classroom observation; teacher self-appraisal; dialogue with the teacher[23]	No	Yes, it is expected to influence professional development activities	Yes, appraisal results can influence decision about promotion and the speed at which a teacher progresses on the salary scale	Yes (to the extent that it allows the teacher to progress on the salary scale)	None	Salary increment withheld; further appraisal; professional development; dismissal; suspension
	Registration	All provisionally registered teachers	Mandatory non-periodic (once, at the end of the conclusion of registration period)	Central education authority or government (The New Zealand Teachers Council is responsible for registering teachers as competent for practice)	Member of school leadership team	National Registration Standards	Planning and preparation; instruction; classroom environment; professional development; links to the community; professional relationships and values; professional leadership; responsiveness to diverse linguistic and cultural backgrounds; analysis and use of assessment information; critical inquiry and problem solving	Classroom observation; teacher self-appraisal; dialogue with the teacher	Yes (2 levels: pass; fail)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decisions about progression to registered teacher status	Yes (to the extent that it allows the teacher to progress on the salary scale)	None	Failure to progress to registered teacher status
Norway	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Poland	Regular appraisal	Teachers in public schools only	Mandatory periodic (annually)	School principal; school board or committee	School principal	A description of the general and professional duties of teachers (as stated in laws and regulations)	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to the community	Classroom observations; may also include teacher self-appraisal and peer evaluation	No	Varies across schools, at the discretion of the school principal and school board or committee, but may inform professional development	No	Varies across schools, at the discretion of the school principal and school board or committee, but may influence the salary	None	None

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Poland (*continued*)	To follow up requests for appraisal by education authorities or stakeholders	All teachers	At the discretion of local education authorities, regional education authorities (education superintendents), the school board, a teacher; the parents council	Central education authority or government (general framework set by law)	School principal	A description of the general and professional duties of teachers (as stated in laws and regulations)	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to the community	Classroom observation; evaluation of documents; formal opinions of senior teachers and student council	Yes (descriptive rating with 3 levels: excellent; good; negative)	Varies across schools, at the discretion of evaluator, but may inform professional development	No	Varies across schools, at the discretion of evaluator, but may influence the salary	None	Dismissal
	Appraisal for career advancement/ promotion[24]	Teachers in public schools only[24]	Voluntary	Central education authority or government (general framework set by law)	School principal	Developmental plan agreed with the school principal	Different aspects of professional performance; subject-area and didactic knowledge	Evaluation of documents (e.g. teacher portfolio); opinion of the parents' council	Yes (2 levels: positive; negative)	No[25]	Yes, appraisal results influence decisions about promotion[26]	Yes (to the extent that it allows the teacher to progress on the salary scale)	None	Negative rating delays career advancement
Portugal	Regular appraisal	All teachers in public schools	Mandatory periodic (every 4 years for permanent teachers) There are 2 components: internal and external appraisal. External appraisal is mandatory only in specific cases (to obtain top rating; at 2 specific career stages; for teachers previously rated "insufficient")	Central education authority; schools	Internal appraisal: teachers from the same school; collegiate body within the school (chaired by principal) External appraisal: trained teachers from other schools	Internal appraisal: school development plan; evaluation parameters established by each school. External appraisal: national evaluation parameters for classroom observation only	Internal appraisal: scientific-pedagogical aspects; participation in school activities and links to the community; professional development External appraisal: classroom observation	Internal appraisal: teacher project (optional); self-appraisal; overall appraisal form used by internal evaluator External appraisal: classroom observation	Yes (5 levels; there is a national quota system for the two top levels)	Yes, it is expected to influence professional development activities (systematic influence for teachers rated "insufficient")	Permanent teachers: appraisal results influence the speed of career progression Non-permanent teachers: appraisal results influence decisions about access to a permanent position and contract renewal	Yes (to the extent that it allows the teacher to progress on the salary scale)	Permanent teachers: extra opportunities for scientific-pedagogical professional development	Permanent teachers: salary increment withheld; further appraisal; compulsory training Non-permanent teachers: permanent contract not granted

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Slovak Republic	Regular appraisal	All teachers	Mandatory periodic (annually)	Central education authority (through the Act on Pedagogical Employees); school principals	School principals	Personal development plan; teacher professional standards	Planning and preparation; instruction; student outcomes; classroom environment; professional development, contribution to school development, links to parents and advisory institutions	Classroom observation; dialogue with the teacher surveys; teacher portfolio	Yes[27]	Varies across schools, it may inform professional development plans	Varies across schools, appraisal results may influence career advancement at both vertical and horizontal levels	Varies across schools, it may influence salary raises	None	Deferral of promotion/career advancement
Slovenia	Regular appraisal	All teachers	Mandatory periodic (annually)	Central education authority; school principal	School principal	School development plan; national regulations on promotion	Planning and preparation; instruction; classroom environment; professional development; links to the community; contribution to school development	Classroom observation; dialogue with the teacher; teacher portfolio	Yes (5 levels)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decisions about promotion and the speed at which a teacher progresses in the career structure	Yes (to the extent that it allows the teacher to progress on the salary scale)	None	Further appraisal
Spain	None	a	a	a	a	a	a	a	a	a	a	a	a	a
Sweden	Registration	All teachers[28]	Mandatory non-periodic (once at the end of the introduction period); in relation to decision on employment status; at the discretion of the school principal; by decision of the Swedish National Agency for Education	Government and central agency (National Agency for Education)	Central agency (National Agency for Education)	National teaching standards	m	Judgement of the school principal	No	No[29]	Yes, appraisal results influence decisions about progression to registered teacher status[30]	No	None	Failure to progress to registered teacher status; withdrawal of registered teacher status through a special board[31]

Table 5.A.1 Teacher appraisal for performance management (2012) (*continued*)

Country	Governance					Procedures				Use of results				
	For what types of performance appraisal is there a framework?	For which teachers does the framework apply?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
United Kingdom (Northern Ireland)	Regular appraisal (Performance Review and Staff Development Scheme, PRSD)	All teachers	Mandatory periodic (annually under the PRSD Scheme)	The Teachers' Negotiating Committee (Employing Authorities, Department and Teacher Unions)	School principal or a teacher reviewer designated by the school principal	Performance Review and Staff Development Scheme (PRSD)	Three personal/shared objectives are set covering the areas of: professional practice; pupil and curriculum development; and personal and professional development	Classroom observation; task observation; review discussion	No, a review statement is prepared	Yes, the PRSD Scheme helps to identify the professional needs and necessary resources to support teachers in their professional development	The PRSD Scheme helps to identify the professional needs and necessary resources to support teachers in their career progression.	The PRSD Review Statement is part of the body of evidence used to inform decisions on pay progression	None	There is an informal stage where a programme of support and development is provided. This may be followed by a formal stage which includes the issue of formal written notice, a targeted support programme and ultimately dismissal if a satisfactory standard of work is not achieved.

Notes: a: information not applicable because the category does not apply; m: information not available.

1. Chile: Only teachers who have been previously rated as "outstanding" or "proficient" can apply to become peer evaluators.

2. Czech Republic: The policy framework is specified through the Act on Pedagogical Employees.

3. France: (1) Civil servants in public schools and salaried employees in private schools: Annually by school principals (ISCED levels 2 and 3) and at irregular intervals by inspectors (on average every 3-4 years at ISCED level 1 and every 6-7 years at ISCED levels 2 and 3). (2) Salaried employees with a permanent contract: every 3 years.

4. Hungary: The Law on Public Employees (*1992. évi XXXIII törvény, 40§*) requires schools to define appraisal procedures.

5. Hungary: According to the general advancement scheme for public employees.

6. Hungary: According to the general advancement scheme for public employees. School leaders have a modest per capita monthly sum to reward teachers' performance. School leaders decide on its distribution among teachers normally for one year. There is no quota, but funds are limited.

7. Israel: Teacher appraisal in Israel has only been introduced recently and at this stage concerns ISCED level 1 and part of ISCED level 2 only. The central policy framework for teacher appraisal for performance management does not apply to ultra-orthodox religious schools.

8. Luxembourg: For further information, see *www.men.public.lu/legislation/lois_rgd_recents/090326_rgd_concours_instituteurs.pdf.*

9. Mexico: *Sistema en Línea para la Evaluación Docente*

10. Mexico: Schemes for the performance management of teachers in schools at ISCED level 3 managed by state education authorities or autonomous agencies are not reflected in the information provided.

11. Mexico: Teachers receive the results of their appraisal in writing (e.g. published on line).

12. Mexico: The Universal Evaluation System (*Evaluación Universal de Docentes*) is being gradually implemented as of 2012, initially covering primary education. It is intended that lower secondary education will be covered in 2013, while pre-primary and special education will be considered in 2014. Private schools will also be covered.

13. Mexico: Teachers can take the examination to apply for a permanent post or to obtain an additional permanent post (permanent posts in Mexico can be defined on an hourly basis and are typically associated with half day duties).

14. Mexico: *Órgano de Evaluación Independiente con carácter federalista.*

15. Mexico: These aspects are defined through the National Exam of Teaching Knowledge and Skills. In some cases, additional aspects are tested.

16. Mexico: *Exámenes Nacionales de Actualización para Maestros en Servicio.*

17. Mexico: *Centro Nacional de Evaluación para la Educación Superior.*

18. Mexico: *Reforma Integral de la Educación Básica.*

19. Mexico: The ENAMS includes 15 standardised multiple-choice tests depending on the teaching area. The number of tests varies depending on national educational priorities.

20. Mexico: Teachers know their test scores as performance feedback. The score may be used in the National Teaching Career Programme (*Programa Nacional de Carrera Magisterial* [PNCM]), see Table 5.A.3.

21: Netherlands: As convened in the terms of employment, made up by central employers in primary education (CAO-PO 9.5.4) or secondary education (CAO-VO 16.2.4).

22: Netherlands: There are central regulations that act as a framework. Within this framework, the school organising bodies (competent authorities) are responsible. The National Council of School Boards acts as a central employer and is in charge of setting the terms of employment.

23. New Zealand: In some cases, information may also be gathered through student surveys and parent surveys.

24. Poland: Teachers wishing to advance on the career ladder take part in this appraisal process.

25. Poland: Professional development is typically part of the developmental plan.

26. Poland: A positive appraisal result is a precondition for career advancement.

27. Slovak Republic: For the rating of teachers, some schools use a descriptive appraisal, some use their own assessment rating scale (excellent, good, satisfactory), or they can use the performance scale recommended by the Ministry (exceptional, very good, standard, partially satisfactory, unsatisfactory)

28. Sweden: This appraisal scheme for the registration of teachers came into force 1 July 2011 as part of the new Education Act. The registration system requires novice teachers to complete an introduction year at a school during which they are supported by a mentor. Upon completion of the introduction year, the school leader is required to assess the teacher as suitable for the profession for the teacher to be registered. The registration is based on the teacher's examination results, a remark by the teacher's principal, and where relevant on additional courses undertaken.

29. Sweden: School principals and school organising bodies may decide on professional development for teachers building upon the registration process (e.g. in new subject areas).

30. Sweden: The central government is planning the development of a multilevel career structure.

31. Sweden: School organising bodies may take additional measures in the case of underperformance.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 5.A.2 Teacher appraisal for completion of probation (2012)

This table describes the appraisal of individual teachers to make a judgement about their work and performance using objective criteria upon completion of a teacher's probationary period. It is, thus, related to a teacher's entry into the profession and designed to evaluate the competence and progress of a newly hired teacher related to the completion of probation.

Country	Governance				Procedures				Use of results				
	For which teachers is there a policy framework for the completion of probation?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13
Australia	All permanent teachers	Mandatory periodic (frequency varies nationally to a maximum of 12 months)	State education authorities or governments; school board or committee	State education authorities; school principal; supervisor; peer evaluator at the same school	State teaching standards; a description of the general and professional duties of teachers; code of conduct From 2013: National teaching standards	Planning and preparation; instruction; classroom environment; professional development; links to the community	Classroom observation; dialogue with evaluators; teacher self-appraisal; teacher portfolio	Yes (2 levels: proficient, not proficient)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decision about access to a permanent position	No	None	Further appraisal; compulsory training; salary increment withheld; permanent contract not granted; failure to pass probationary period; dismissal
Austria	No	a	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	No	A	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fr.)	No	a	a	a	a	a	a	a	a	a	a	a	a
Canada	Teachers during probation	In relation to decision on employment status	Provincial/territorial education authorities or governments	School principal; superintendent	Provincial teacher standards or competencies	Instruction (engages students; applies creativity and innovation); classroom environment (inclusive learning environment); professional responsibilities	Classroom observation; dialogue with the teacher; peer collaboration; parent survey	Varies across provinces/territories (e.g. 2 levels: pass, fail; performing in a satisfactory manner, performing in a non-satisfactory manner)	Yes, it is expected to influence professional development activities	Yes	No	Recognition; professional learning; employment opportunities	Compulsory training; contract not renewed; permanent contract not granted; loss of certification

Table 5.A.2 Teacher appraisal for completion of probation (2012) (*continued*)

Country	Governance				Procedures				Use of results				
	For which teachers is there a policy framework for the completion of probation?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13
Canada (*continued*)	All teachers at the end of probation and new hires during first year (except in the province of Quebec)	Mandatory periodic (at the end of probationary period); new hires formally appraised twice during the first year	Provincial/territorial education authorities or governments	School principal	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development; contribution to school development	Classroom observation; dialogue with evaluators; peer collaboration; parent survey	Varies across provinces/territories (e.g. satisfactory, development needed, unsatisfactory)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence the speed at which a teacher progresses in the career structure	No	None	Failure to pass probationary period; permanent contract not granted; withdrawal or inaccessibility to the priority list
Chile	No	a	a	a	a	a	a	a	a	a	a	a	a
Czech Republic	No	a	a	a	a	a	a	a	a	a	a	a	a
Denmark	No	a	a	a	a	a	a	a	a	a	a	a	a
Estonia	No	a	a	a	a	a	a	a	a	a	a	a	a
Finland	No	a	a	a	a	a	a	a	a	a	a	a	a
France	All teachers	In relation to decision on employment status at the end of the probationary period (1 year)	Central education authority (by ministerial order)	General and Local Inspectorates; school principal (ISCED 2, 3); supervisor	National norms and standards (competency framework in form of a ministerial order)	Planning and preparation; instruction; classroom environment; professional development (includes capacity for innovation) contribution to school development; links to the community (parents and school partners)	Report of supervisor; opinion of school principal (ISCED 2, 3); inspector's report	Yes (ISCED level 1: range of scores and descriptive ratings depending on the authority [e.g. poor, fair]; ISCED levels 2 and 3: range of scores)	No	Yes, appraisal results influence the speed at which a teacher progresses in the career structure and salary scale; appraisal results also influence decisions about a teacher's access to a permanent position	Yes (to the extent that it allows the teacher to progress on the career structure and salary scale)	None	Dismissal; extension of probationary period (1 year); relegation to previous status or post
Hungary	No	a	a	a	a	a	a	a	a	a	a	a	a
Iceland	No	a	a	a	a	a	a	a	a	a	a	a	a

Table 5.A.2 Teacher appraisal for completion of probation (2012) (*continued*)

Country	Governance				Procedures				Use of results				
	For which teachers is there a policy framework for the completion of probation?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13
Ireland	All teachers	ISCED level 1: twice during the probationary period ISCED levels 2 and 3: at least three months before the end of the probationary period (1 year)	Central education authority in consultation with the teacher professional organisation (Teaching Council)	ISCED level 1: Inspectorate of the Department of Education and Skills ISCED levels 2 and 3: school leadership	ISCED 1: evaluation criteria for probation published by the Inspectorate; assessment template which provides for ratings in relation to main aspects of practice. ISCED 2: none[1]	Planning and preparation; instruction; classroom environment	Classroom observation; dialogue with the teacher	ISCED level 1: Yes (3 levels: pass; fail; extension of probationary period[2]) ISCED levels 2 and 3: Yes (2 levels: pass; fail)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence the speed at which a teacher progresses in the career structure	No	Some categories of teaching posts are only open to teachers who have successfully completed probation	ISCED level 1: failure to pass probationary period as determined by inspector; further appraisal; teacher does not achieve full registration with the Teaching Council; implications for a teacher's eligibility for employment in a school ISCED levels 2 and 3: failure to pass probationary period within 3-year period as certified by school principal; teacher does not achieve full registration with the Teaching Council; implications for a teacher's eligibility for employment in a school
Israel	Teachers in public and government-dependent private schools only[3]	At the end of the probationary period (2-3 years)	Central education authority	School principal; central education authority	National teaching standards	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to the community	Classroom observation; dialogue with the teacher; teacher self-appraisal; teacher portfolio	Yes (5 levels)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decision about access to a permanent position	No	None (under discussion)	Further appraisal; failure to pass probationary period; dismissal
Italy	Permanent teachers in public schools only	At the end of the probationary period (1 year)	Central education authority	School principal; school-based Evaluation Committee[4]	None[5]	Professional development; presence at school (minimum 180 days); discussion of written essay; other aspects (not specified)[6]	Not specified (criteria are determined by individual school leaders and evaluation committees)	Yes (2 levels: pass; fail)	No	Yes, appraisal results influence decision about access to a permanent position	No	None	Failure to pass probationary period; dismissal[7]
Korea	No	a	a	a	a	a	a	a	a	a	a	a	a

Table 5.A.2 Teacher appraisal for completion of probation (2012) (*continued*)

Country	Governance				Procedures				Use of results				
	For which teachers is there a policy framework for the completion of probation?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13
Luxembourg	All teachers at ISCED levels 2 and 3 only[8]	At the end of the probationary period (2 years)	Central education authority	School principal; teacher education faculty of University of Luxembourg	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development	Teacher testing (as part of the national recruitment exam), dialogue with the teacher, classroom observation, teacher portfolio, dissertation	Yes (2: pass; fail)	No	Yes, appraisal results influence decisions about the terms of the contract	No	None	Possibility to re-take the examination; failure to pass probationary period; contract not granted
Mexico	No	a	a	a	a	a	a	a	a	a	a	a	a
Netherlands	All teachers	At the end of the probationary period	Central employer (National Council of School Boards); school organising bodies (competent authorities)[9]	School principal representing school organising body (competent authority)	National teaching standards	Organisational; pedagogical and subject matter competences; interpersonal competences; teamwork; links to the community; professional development	Extensive descriptions of competencies	Varies across schools depending on school internal regulations	Yes, it is expected to result in a professional development plan	Varies across schools depending on school internal regulations	Varies across schools depending on school internal regulations	Changes in work responsibilities; extra opportunities for in-service professional development	Dismissal
New Zealand	All newly trained teachers prior to teacher registration	At the end of the probationary period (duration varies according to hours worked)	Teacher professional organisation (The New Zealand Teachers Council)	School principal	National registration standards	Professional relationships and values; professional knowledge and practice	Classroom observation; discussion; documentation	Only a decision about teacher registration	Yes, it is expected to influence professional development activities	Teachers need to achieve teacher registration to continue in teaching career	No	They can become registered teachers	Provisionally registered teachers who are not successful can take more time and try again
Norway	No	a	a	a	a	a	a	a	a	a	a	a	a
Poland	No	a	a	a	a	a	a	a	a	a	a	a	a
Portugal	All teachers in public schools	At the end of the probationary period	Central education authority; schools	Teachers from the same school; collegiate body within the school (chaired by principal); trained teachers from other schools	School development plan; evaluation parameters established by each school; national evaluation parameters for classroom observation only	Instruction; scientific-pedagogical aspects; participation in school activities and links to the community; professional development	Teacher project (optional); self-appraisal; overall appraisal form used by internal evaluator; classroom observation	Yes (5 levels; there is a national quota system for the two top levels)	Yes, it is expected to influence professional development activities (systematic influence for teachers rated "insufficient")	m	m	m	m

Table 5.A.2 Teacher appraisal for completion of probation (2012) (*continued*)

Country	Governance				Procedures				Use of results				
	For which teachers is there a policy framework for the completion of probation?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Does the appraisal inform the teacher's professional development activities?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?	What are the responses to underperformance of teachers?
	1	2	3	4	5	6	7	8	9	10	11	12	13
Slovak Republic	All beginner teachers	Pursuant to the legislation, the periodicity of teacher appraisal is set individually in each school (probationary period typically lasts 1 year)	Central education authority (through Act on Pedagogical Employees; Decree of the Ministry of Education)	The mentor teacher and the examination committee nominated by the school principal	Plan for adaptation education; teacher professional standards	Planning and preparation; instruction; classroom environment; professional development, contribution to school development, links to parents and advisory institutions	Classroom observation; dialogue with the teacher surveys; teacher portfolio; student outcomes	Yes[10]	Varies across schools depending on individual needs and capabilities, it may inform teacher professional development plans	Yes, appraisal results influence decisions about promotion of the teacher to the next career level (independent teacher)	Varies across schools, appraisal results may influence career advancement	None	Failure to pass probationary period; access to the next career level not granted
Slovenia	All teachers	At the end of the probationary period	Central education authority	Central education authority	None	Pedagogical and subject knowledge; language competency; knowledge of legislation	Teacher testing	Yes (2 levels: pass; fail)	No	No[11]	No	None	Failure to pass probationary period
Spain	No	a	a	a	a	a	a	a	a	a	a	a	a
Sweden	All teachers	At the end of the probationary period (1 year)	Central education authority and government	Central education authority; school principal	National teaching standards	Planning and preparation; instruction; classroom environment; professional development; contribution to school development; links to community	School principal's judgement informs application for registration with the Swedish National Agency for Education	No	No	Yes, appraisal results influence decisions about access to a permanent position or a fixed term contract	No	None	Teacher cannot apply for registration; possibility to re-start a probationary period
United Kingdom (Northern Ireland)	All teachers during Induction and Early Professional Development (EPD)	During Induction and EPD under the Performance Review and Staff Development (PRSD) Scheme (It is not mandatory for teachers to complete Induction and EPD to be registered but it is normal practice)	The Teachers' Negotiating Committee (Employing Authorities, Department and Teacher Unions)	School principal or a teacher reviewer designated by the school principal	Performance Review and Staff Development Scheme (PRSD)	Three personal/shared objectives are set covering the areas of: professional practice; pupil and curriculum development; and personal and professional development	Classroom observation; task observation; review discussion	No, a review statement is prepared	Yes, the PRSD Scheme helps to identify the professional needs and necessary resources to support teachers in their professional development	The PRSD Scheme helps to identify the professional needs and necessary resources to support teachers in their career progression.	The PRSD Review Statement is part of the body of evidence used to inform decisions on pay progression	None	There is an informal stage where a programme of support and development is provided. This may be followed by a formal stage which includes the issue of formal written notice, a targeted support programme and ultimately dismissal if a satisfactory standard of work is not achieved.

Notes: a: information not applicable because the category does not apply; m: information not available

1. Ireland: A code of practice is being developed by the Teaching Council and is currently open for consultation.

2. Ireland: Usually used only once.

3. Israel: Teacher appraisal has only been introduced recently and at this stage concerns ISCED level 1 and part of ISCED level 2 only. The central policy framework for teacher appraisal for the completion of probation does not apply to ultra-orthodox religious schools.

4. Italy: Completion of probation regulated with Legislative Decree n. 297/1994, artt. 11, 438, 439, 440 and Ministerial Newsletter 196/2006. The Evaluation Committee comprises on average four teachers and the school leader who functions as president of the committee.

5. Italy: Beyond basic central requirements regarding participation in training, presence at school and discussion of a written essay, appraisal criteria are at the discretion of the school principal.

6. Italy: Further aspects are determined by the school principal.

7. Italy: Failure to pass the probationary period requires a teacher to repeat the probationary period. In case of a second unsatisfactory performance the teacher can be dismissed, return to the original incoming institution (very few cases) or can have an additional year to complete the evaluation process (Legislative Decree 297/1994, art. 439). Non completion of probation implies that the teacher does not have access to a permanent position.

8. Luxembourg: Teachers must have passed the national recruitment examination to be admitted to a teacher probationary period. For further information on this appraisal process see *www.men.public.lu/sys_edu/personnel_ecoles/090326_recrutement_prof_postprimaire/index.html*.

9. Netherlands: There are central regulations that act as a framework. Within this framework, the school organising bodies (competent authorities) are responsible. The National Council of School Boards acts as a central employer and is in charge of setting the terms of employment.

10. Slovak Republic: For the rating of teachers, some schools use a descriptive appraisal, some use their own assessment rating scale (excellent, good, satisfactory), or they can use the performance scale recommended by the Ministry (exceptional, very good, standard, partially satisfactory, unsatisfactory)

11. Slovenia: Teachers that are judged as having failed their probationary period are not granted a permanent position.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 5.A.3 Teacher appraisal for rewards (2012)

This table describes the appraisal of individual teachers to make a judgement about their work and performance using objective criteria for rewards purposes, i.e. appraisal schemes that are exclusively designed with the objective of providing rewards to teachers. Performance management schemes which may also lead to rewards, but are not exclusively designed as rewards schemes are included in Table 5.A.1.

Country	Governance				Procedures				Use of results		
	For which teachers is there a policy framework for a rewards scheme?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?
	1	2	3	4	5	6	7	8	9	10	11
Chile	Teachers in public and publicly subsidised private schools (Pedagogical Excellence Allowance [AEP])[1]	Voluntary	Central education authority	Central education authority	National teaching standards	Knowledge of discipline and curricular content; pedagogical knowledge; skills and competencies in the classroom	Teacher testing; teacher portfolio	No	No	Yes, a pay allowance is provided for good performance (10 years, no quota)	None
	Teachers in public schools who obtain either of the two top scores in the regular teacher appraisal system (Variable Individual Performance Allowance [AVDI])[2]	Voluntary	Central education authority	Central education authority	National teaching standards	Knowledge of discipline and curricular content; pedagogical knowledge; skills and competencies in the classroom	Teacher testing	Yes (3 levels)	No	Yes, a pay allowance is provided for good performance (2 to 4 years – until next regular teacher appraisal, no quota)	None
	Groups of teachers (teaching bodies of individual schools) in public and publicly subsidised private schools (National Performance Evaluation System [SNED])[3]	Mandatory (annually)	Central education authority	Central education authority	Student learning objectives; school development plan	Aspects are assessed at the school level: student performance; ability to innovate; equality of opportunities; links to school community	Student outcomes (standardised assessment results); variety of school indicators (e.g. student retention rates; enrolment of students with special needs) school development plan; school policies	It results in a rating for the school	No	Yes, schools within top 35th percentile receive extra subsidy (Subsidy for Performance of Excellence), which is distributed among their teaching bodies (in proportion to contract hours; the school can distribute 10% of subsidy to teachers according to its own criteria)	None

Table 5.A.3 Teacher appraisal for rewards (2012) (*continued*)

Country	Governance				Procedures				Use of results		
	For which teachers is there a policy framework for a rewards scheme?	Under which circumstances are teachers appraised, and how often?	Who determines the procedures for teacher appraisal?	Who are the evaluators?	Against what references are teachers appraised?	What aspects of teacher performance are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the teacher?	Do appraisal results impact career advancement?	Do appraisal results impact pay levels?	What other rewards may teacher appraisal involve?
	1	2	3	4	5	6	7	8	9	10	11
Korea	All teachers (Performance-based Incentive System)	Mandatory periodic (annually)	Central education authority	School principal; peer evaluators at the same school	A description of the general and professional duties of teachers	Planning and preparation; instruction; classroom environment; professional development; links to the community; student guidance	Observation of performance	Yes (3 levels)	No	Yes, a pay allowance is provided for good performance (once)	None
Mexico	Teachers in public schools only (National Teaching Career Programme [PNCM])[4]	Voluntary (once per year)	Central education authority or government (Secretariat of Public Education [SEP])[5] and Teacher Union (SNTE)[6] (through the National Academic Commission)	Central education authority or government (Secretariat of Public Education [SEP])[5]	A description of the general and professional duties of teachers; school development plan or school project	Planning and preparation; instruction; professional development; contribution to school development; links to the community; years of teaching service	Teacher testing; student outcomes (e.g. standardised assessment results; graduation, retention, or pass rates)	No	No[7]	Yes, appraisal results affect salary allowances[7]	None

Notes:

1. Chile: *Asignación de Excelencia Pedagógica.*

2. Chile: *Asignación Variable por Desempeño Individual.*

3. Chile: *Sistema Nacional de Evaluación de Desempeño.*

4. Mexico: *Programa Nacional de Carrera Magisterial.*

5. Mexico: *Secretaría de Educación Pública.*

6. Mexico: *Sindicato Nacional de Trabajadores de la Educación.*

7. Mexico: The appraisal results grant access to a system of salary allowances with five different levels not associated with vertical differentiation of the teacher's tasks.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

A corrigendum has been issued for this page. See: http://www.oecd.org/about/publishing/Corrigendum_Synergies-for-better-learning.pdf

Table 5.A.4 Employment status and career development of teachers (2012)

This table describes the employment status and career development of teachers.

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
Australia	State education authorities or governments; local education authorities; school, school board or committee[1]	Civil servant status; salaried employee status	Yes, both teachers with civil servant status and salaried employee status (maximum period of time ranges from 1 to 5 years nationally)	Multilevel career structure, with a salary scale for each career level (number of levels varies nationally)	Length of service; completion of professional development; taking on extra roles and tasks; teacher appraisal results
Austria	ISCED level 1 (public schools): state education authorities ISCED level 2 (public schools): central or state education authorities depending on the type of schools ISCED level 3 (public schools): central education authority	Civil servant status; salaried employee status	Yes, teachers with salaried employee status (for a maximum of 5 years)	General: unique career stage with a single salary scale Pre-voc and voc: multilevel career structure with a single salary scale (2 career levels)	General: salary step increments based on length of service (biennial progress) Pre-voc and voc: salary step increments based on length of service (biennial progress); taking on extra roles and tasks
Belgium (Fl.)	School organising bodies	Civil servant status	Yes	Unique career stage with a single salary scale[2]	Salary step increments based on length of service
Belgium (Fr.)	School organising bodies	Civil servant status; salaried employee status[3]	Yes, teachers with salaried employee status	Unique career stage with a single salary scale	Salary step increments based on length of service
Canada	Provincial/territorial education authorities or governments; school, school board, or committee	Salaried employee status	Yes (maximum period of time varies across provinces/territories and is at the discretion of education authorities depending on the availability of permanent teaching positions)	Multilevel career structure (number of levels varies; 10, 11, or 12 steps in the salary scale)	Length of service; teacher appraisal results; extra roles and tasks

A corrigendum has been issued for this page. See: http://www.oecd.org/about/publishing/Corrigendum_Synergies-for-better-learning.pdf

Table 5.A.4 Employment status and career development of teachers (2012) (*continued*)

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
Chile	School organising bodies (sustainers)	Public schools: salaried employee status Government-dependent private schools: salaried employee status	Public schools: Yes, up to 20% of the teachers with salaried employee status Government-dependent private schools: Yes, up to 100% of teachers with salaried employee status	Public schools: unique career stage within a single salary scale[4] Government-dependent private schools: at the discretion of the school organising bodies (sustainers) to determine career structure and salary scale beyond base salary	Public institutions: salary step increments based on length of service Government-dependent private institutions: at the discretion of the school organising bodies (sustainers)
Czech Republic	School	Public schools: civil servant status Private schools: salaried employee status	Yes, both teachers with civil servant and salaried employee status (twice in a row)	Multilevel career structure within a salary scale for each career level (3 career levels)	Length of service; completion of professional development; teacher appraisal results; extra roles and tasks
Denmark	ISCED levels 1 and 2 (public schools): local education authorities ISCED levels 1 and 2 (private independent schools): school organising bodies ISCED level 3 (all schools): school board	Salaried employee status	No[5]	None	a
Estonia	School	Salaried employee status	Yes[6]	Multilevel career structure with a salary scale for each career level (4 career levels)	Extra roles and tasks; completion of professional development
Finland	Local education authorities	Civil servant status; salaried employee status	No[7]	Unique career stage with a single salary scale	Salary step increments based on length of service; taking on extra roles and tasks (e.g. deputy director)

Table 5.A.4 Employment status and career development of teachers (2012) (*continued*)

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
France	Central education authority (Ministry of Education, for civil servants); regional education authorities (*recteurs d'académies*; for contract (public or private) employees [*contractuels*]), schools (for replacement teachers [*vacataires*]); superior in addition to central education authority (Ministry of Education) and the school (for contractors in private schools)	Civil servant status; public contract employees (*contractuels de droit public*); replacement teacher status (*vacataires)*; salaried employee status (private law with a simple contract)	Yes, replacement teachers for a maximum of 6 years with a temporary contract after which they may receive a permanent contract)	Multilevel career structure with a salary scale for each career level (2 career levels with various corresponding salary steps; depending on the teaching body each level corresponds to one base salary index)	Length of service; teacher appraisal results; completion of professional development for promotion to a higher level in the career structure
Hungary	School (until 2013); Klebelsberg Institution Maintenance Centre (from 2013)	Public schools: Public employee Private schools: Salaried employee	Yes (for a maximum of 1 year)	Multilevel career structure with a salary scale for each career level (5 career levels defined by qualification, 14 salary steps within each career level, progression on the salary scale every 3 years)	Length of service; completion of professional development; taking on extra roles and tasks; teacher appraisal results
Iceland	ISCED levels 1 and 2: local education authorities ISCED level 3: central education authority	ISCED levels 1 and 2: salaried employee status ISCED level 3: civil servant status	Yes (for a maximum of 2 years)	ISCED levels 1 and 2: multilevel career structure (3 levels with nine steps in the salary scale) ISCED level 3: unique career stage with a single salary scale (9 steps)	Length of service; taking on extra roles and tasks; completion of professional development; age; administrative responsibilities
Ireland	School board of management or school manager or local education authority[8]	Salaried employee status; state non-civil service status (public servant)	Yes	Unique career stage with a single salary scale (17 steps)	Salary step increments based on length of service; taking on extra roles and tasks

Table 5.A.4 Employment status and career development of teachers (2012) (*continued*)

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
Israel	ISCED levels 1 and 2: central education authority; corporation or non-profit organisation ISCED level 3: local education authorities; corporation or non-profit organisation	ISCED levels 1, 2 and 3 (all schools): civil servant status; salaried employee status	ISCED levels 1, 2 and 3 (all schools): Yes, both teachers with civil servant status and salaried employee status (Generally for a maximum of 2-3 years)	ISCED level 1 (public and government-dependent private schools only): multilevel career structure with a salary scale for each career level (3 levels: 9 steps at level 1; 4 steps at levels 2 and 3) ISCED level 2 (public and government-dependent private schools only): multilevel career structure with a salary scale for each career level (4 levels: 9 steps at levels 1 and 2; 4 steps at level 3 and 4) ISCED level 3 (public and government-dependent private schools only): multilevel career structure with a single salary scale (4 levels)	ISCED levels 1 and 2 (public and government-dependent private schools only): length of service; completion of professional development; teacher appraisal results ISCED level 3 (public and government-dependent private schools only): length of service; completion of professional development
Italy	Central education authority (public schools only)[9]	Civil servant status	Yes	Multilevel career structure with a salary scale for each career level (4 levels with 21 salary steps each)	Length of service
Korea	Provincial/regional education authorities	Civil servant status	No	Multilevel career structure with a single salary scale (3 levels with 50 steps in a single salary scale)	Length of service; completion of professional development
Luxembourg	Central education authority or government (public schools)	Public schools: civil servant status	No	Unique career stage with a single salary scale	Salary step increments based on length of service
Mexico	State education authorities (public schools)	Salaried employee status	Yes	Unique career stage with a single salary scale (and 5 levels of salary allowances depending on voluntary teacher appraisal)	Length of service; completion of professional development; teacher appraisal results to access salary allowances

Table 5.A.4 Employment status and career development of teachers (2012) (*continued*)

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
Netherlands	Public schools: municipality or bodies with powers transferred by municipality Private schools: school organising bodies (competent authorities)	Public schools: civil servant status Private schools: salaried employee status[10]	Yes (for a maximum of 3 years)	ISCED level 1: multilevel career structure with a salary scale for each career level (2 levels with 15 steps in the salary scale each) ISCED levels 2 and 3: multilevel career structure (3 levels with 15 steps in the salary scale each)	Salary step increments based on teacher appraisal results; taking on extra roles and tasks
New Zealand	School Board of Trustees	Salaried employee status	Yes[11]	Unique career stage with a single salary scale (14 steps within the salary scale)[12]	Salary step increments based on length of service; taking on extra roles and tasks
Norway	Local education authorities	Salaried employee status	Yes[13]	Multilevel career structure with a single salary scale (5 steps within the salary scale)[14]	Length of service; taking on extra roles and tasks
Poland	School	Trainee and contract teachers: employment agreementAppointed and chartered teachers: employment based on appointment	Yes, trainee teachers and contract teachers when substituting an absent teacher	Multilevel career structure with a salary scale for each career level (4 levels, each with corresponding scale of basic salary: trainee (probation), contract, appointed, and chartered)	Length of service; teacher appraisal results; qualifications; interviews; examinations
Portugal	Central education authority	State civil servant status or employee with fixed contractual status	Yes	Unique career stage with ten steps in the salary scale	Length of service; teacher appraisal results; completion of professional development (stricter requirements to progress to the 5th and 7th step of the salary scale)
Slovak Republic	School and/or school organising bodies (in case of schools that do not have a legal personality)	Civil servant status	Yes	Multilevel career structure with a single salary scale (beginner teacher, independent teacher, teacher with 1st certification level, teacher with 2nd certification level)	Qualifications, length of service; teacher appraisal results

Table 5.A.4 Employment status and career development of teachers (2012) (*continued*)

Country	Employment status			Career development	
	Who is the employer of teachers?	What is the employment status of teachers?	Can teachers be employed on fixed-term contracts?	What is the structure of the teaching career?	What determines teacher career progression?
	1	2	3	4	5
Slovenia	School	Civil servant status	Yes (for a maximum of 2 years)	Multilevel career structure (3 levels: mentor, advisor, counsellor)	Length of service; completion of professional development; taking on extra roles and tasks
Spain	State education authorities or governments	Civil servant status; salaried employee status	Yes, teachers with salaried employee status (for a maximum of 1 year)[15]	Multilevel career structure with a single salary scale (5 steps within the salary scale)	Length of service; taking on extra roles and tasks[16]
Sweden	School organising bodies (local education authorities or independent schools)	Salaried employee status[17]	Yes, teachers with salaried employee status on the decision of the local board	Unique career stage with a single salary scale[18]	Salary step increments based on length of service; completion of professional development
United Kingdom (Northern Ireland)	The relevant Employing Authority – can be one of five Education and Library Boards, CCMS or Boards of Governors of Voluntary Grammar and Grant-Maintained Integrated Schools. However, with the implementation of the Education and Skills Authority in 2013, ESA will become the employer for all teachers in grant-aided schools.	Public servant, salaried employee status	Yes	Multilevel career structure, with a salary scale for each career level	Length of service; taking on extra roles and tasks; teacher appraisal results

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational.

1. Australia: Australia is a federation of eight states and territories. There are differences in employment practices between states and territories, as well as differences between public (government) and private (non-government) schools. In private schools that are part of a system teachers are often appointed by the local education authority (system). In private schools that are not part of a system, teachers are appointed by the school, school board or committee.

2. Belgium (Fl.): The career structure is linked to educational levels (ISCED levels 1, 2 and 3), educational stages at ISCED levels 2 and 3, the educational programme and the subjects taught at ISCED levels 2 and 3 (general, pre-voc and voc) and teacher qualifications. Differences in salary are mainly related to differences in qualifications (different qualifications are required for teaching at different ISCED levels). The legal status of teachers in the Flemish Community of Belgium is established by Decrees. Notable characteristics are: (1) A teacher's career generally starts with a few years of supply teaching or temporary teaching. The teacher's status at this stage is referred to as "temporary appointment for a limited period of time". (2) After a minimum of 720 days of seniority during which a teacher has worked 600 days effectively in the position concerned, a teacher can ask for a "temporary appointment for a continuous period of time" (*Tijdelijke aanstelling van*

doorlopende duur [TADD]). This status gives the teacher more job security. Moreover, teachers can now apply for a "permanent appointment", which holds even more social benefits and job security. (3) In order to get a permanent appointment teachers are required to meet a number of conditions. Apart from general legal requirements (e.g. nationality, language competency, certificate of good conduct) these are: The teacher must have the official qualifications for the job (e.g. diploma, certificate). On 30 June of the school year before a teacher is to be permanently appointed a teacher must have 720 days of seniority of which he or she has worked 360 days effectively in the position concerned; a teacher's last evaluation report, if available, must be positive; on 31 December before a teacher is to be permanently appointed the teacher must have a "temporary appointment for a continuous period of time" for the job the teacher is to be permanently appointed to; and Teaching must be the teacher's main profession.

3. Belgium (Fr.): Teachers are employed as salaried employees until they receive civil servant status.

4. Chile: Biennial rates up to 100% of base salary and training recognised until 40% of base salary.

5. Denmark: In certain cases it is possible to employ teachers on fixed-term contracts. A specific reason is required in this case (e.g. that the teacher replaces a permanent teacher who is absent due to illness or leave).

6. Estonia: If in a competition organised for filling a vacant teacher position, none of the applicants meets the qualification requirements, the school principal may hire a person that has completed at least secondary education on a fixed-term employment contract for a period of up to one year.

7. Finland: Teachers are appointed until retirement.

8. Ireland: Teacher salaries are paid by the central education authority.

9. Italy: The national labour contracts for teachers (CCNL) dated 2002-05 and 2006-09 regulate the contracts of teachers hired by state schools only, not those hired by regions, by provinces or municipalities. Teachers are considered civil servants also according to legislative decree 150/2009 (Legge Brunetta). In addition, the contractual arrangements for all civil servants in Italy are centralised within a specific agency (ARAN) dealing specifically with public employment.

10. Netherlands: The terms of employment for teachers as civil servants and salaried employees are identical.

11. New Zealand: The employment of teachers is subject to the provisions of the Employment Relations Act in relation to fixed term employment generally (e.g. will end on a specified date, or on the occurrence of a specified event, or conclusion of a specified project). The category of relieving teachers is frequently used for fixed term employees.

12. New Zealand: Additional rewards on a fixed-term or permanent basis for a range of management responsibilities.

13. Norway: If no applicant satisfies the qualification requirements laid down in the Education Act, a temporary appointment may be made. Unless a shorter period of appointment is agreed, such appointments shall last until 31 July.

14. Norway: There are also some local positions with locally negotiated salaries.

15. Spain: Most teachers in public schools are civil servants. A small percentage of public school teachers, however, have a salaried employee status (*profesores interinos*) and are employed to teach a single academic course. Teachers in government-dependent private institutions and in independent private institutions are employed according to employment legislation, sometimes on a single year basis and sometimes on a long-term basis.

16. Spain: School principals and other members of the school leadership team have been included as they are teachers and keep direct teaching obligations. Except for these leadership positions and other extra roles and tasks, salary increments depend mostly on the length of service.

17. Sweden: Teachers in Sami schools and special schools have civil servants status.

18. Sweden: The central government is planning the development of a multilevel career structure.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Chapter 6

School evaluation:

From compliancy to quality

School evaluation plays an important role in the evaluation and assessment framework and can exert considerable influence. This chapter presents evidence on different approaches to external school evaluation, school self-evaluation and the use of comparative school performance measures. It examines governance issues, different procedures used, the capacity for undertaking and using the results of school evaluation and the reporting of results. It then presents some options seeking to promote a balance of policies to better serve school improvement.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter analyses approaches to school evaluation within the evaluation and assessment framework. School evaluation refers to the evaluation of individual schools as organisations. This chapter covers internal school evaluation (school self-evaluation or review), external school evaluation (e.g. school reviews, school inspections) and the comparison of schools on different performance measures.

School evaluation is increasingly considered as a potential lever of change that could assist with decision making, resource allocation and school improvement, especially as: further autonomy is given to individual schools, market forms of accountability gain in importance, and the school is increasingly recognised as the key agency within the education system for improving student learning.

The effective monitoring and evaluation of schools is central to the continuous improvement of student learning: Schools need feedback on their performance to help them identify how to improve their practices; and schools should be accountable for their performance.

This chapter is organised in eight sections. After this introduction, the second section lays out the analytical approach, followed by a third section on impact, drivers and contextual developments. The following four sections describe key features of school evaluation and country practices, structured along the four main topics of the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: governance, procedures, capacity and use of results. The final section provides pointers for policy development.

Analytical approach

Scope and definitions

School evaluation concentrates on key processes such as teaching and learning, school leadership, educational administration, school environment and the management of human resources. It does so in association with an analysis of student outcomes, both the achievement/progress of students and the equity of student results. It also takes into account inputs such as the infrastructure, funding and characteristics of the school staff. This report defines school evaluation as an evaluation of the following major aspects:

- the effectiveness of the structures and processes in place within a school
- the implementation of national educational policies and regulations within the school
- the quality of student learning outcomes at the school
- the capacity for schools to improve.

This chapter examines three major approaches to school evaluation:

- **School self-evaluation or review**: This concerns an evaluation or review conducted by members of the school to assess the effectiveness of structures and processes in place and the quality of student learning outcomes. Such internal reviews of school effectiveness and quality may draw on input from school leadership, teachers, other staff, students, parents and the school community.

- **External school evaluation or review**: This concerns the evaluation or review of the quality of structures and processes operating within a school and the quality of student learning outcomes as judged by an external body. External reviews may be conducted by specific national or state institutions, such as Inspectorates or Quality Review Agencies, by a group of officials within a government department or Ministry of Education or by accredited individuals. In these cases, external review typically involves a strong focus on accountability, but increasingly aims to give feedback for school development. External reviews may also be conducted by professionals in other schools in the nature of "collegial" or "peer" reviews. In these cases, external review typically focuses on school improvement and can be taken up by schools as part of their own self-evaluation activities.
- **The comparison of schools on different performance measures**: This typically aims to compare schools on standardised measures to allow the benchmarking of their performance in relation to other schools, particular districts or regions or national averages. Such comparative performance measures may be reported to schools for internal use in their own evaluation processes and/or may be reported publicly to allow a wider audience to compare schools. The argument for the latter is generally linked to providing parents and students with information on which to base decisions of school choice.

Conceptual framework

The OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes uses a conceptual framework to summarise the aspects involved in school evaluation and the way these interconnect (see Figure 6.1). The overarching policy objective is to ensure that school evaluation contributes to the improvement of student outcomes through improved school practices. There is a complex range of features associated with school evaluation. This chapter presents these in four major areas:

- **Governing school evaluation**: This addresses the purpose of school evaluation and includes the major responsibilities for devising and conducting school evaluation and setting a legal framework for school evaluation. It also refers to how external school evaluation and school self-evaluation are articulated.
- **Procedures used in school evaluation**: This aspect refers to the features of a given approach to school evaluation, that is, the mix of instruments, criteria and standards, knowledge and skills used in a specific school evaluation model. It also includes decisions about the population of schools involved, the reference standards, the character of the evaluation, the nature of externality, the steps of the process, and the frequency.
- **Competencies to evaluate schools and to use the results of school evaluation**: This aspect concerns the preparation to evaluate, to be evaluated and to use the results of an evaluation as well as the choice of the groups undertaking these functions. It includes issues such as: the choice of the evaluators and the development of the skills to perform the evaluation of a school; the preparation by schools to be the subject of an evaluation; the development of competencies to effectively use the results of an evaluation for the improvement of school practices; and the design of agencies to review school evaluation results with a view to hold schools accountable and to inform policy development.

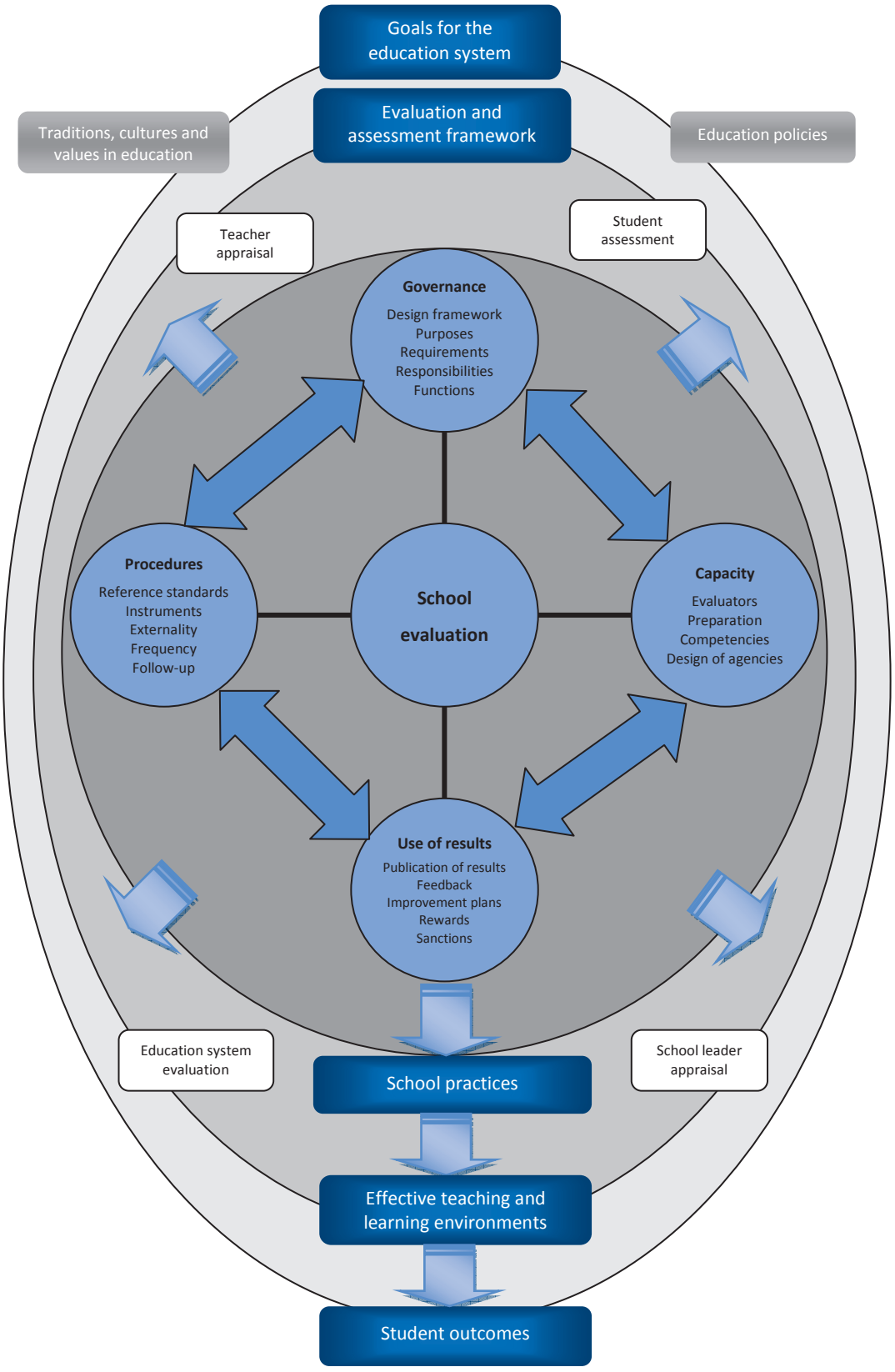
- **Using the results of school evaluation**: This encompasses the objectives of a particular school evaluation process and the mechanisms designed to ensure that evaluation results are used in a way such objectives are reached. The objectives of school evaluation typically consist of feedback for improvement, accountability for performance and information about the quality of school practices. Examples of mechanisms to use evaluation results include feedback and recommendations for improvement, an improvement plan, publication of school-level results, financial and other rewards as well as sanctions.

Impact, drivers and contextual developments

School evaluation policies, like all components of the evaluation and assessment framework, have been influenced and shaped by wider trends in public management (see Chapter 2). With devolved responsibilities, there are greater demands to hold schools accountable for their quality. This means a greater level of responsibility at the school level for quality improvement; a greater focus on the outcomes a school secures for its students; and demands for the public to have access to information on school quality. For example, in Mexico the National Model for Total Quality in Mexico, which was drawn up to promote a general management approach for quality assurance in public services, led to the development of a voluntary System for School Self-Evaluation for Quality Management in 2007 (SEP and INEE, 2011). In the Flemish government, trends for greater transparency with the "Active publicity" policy led to the publication of inspection reports for individual schools on the Inspectorate's website from 2007 (Flemish Ministry of Education and Training, 2010).

Perhaps the most specific external policy influence on school evaluation has been the wider policy trends on the approach to external evaluations in the public sector. This has impacted the approach to external school evaluations significantly. There is an increasing focus on a need to maximise the benefits of external school evaluation activities, but to minimise the potential burden that these may place on school time. This is often in the larger context of public sector reform to place more emphasis on outcomes and impact, coupled with robust self-evaluation and a reduced, more proportionate approach to external supervision. It is also fundamentally linked to a concern to make more effective use of the resources available for external evaluation. For example, in the Netherlands there is a programme to reform national inspections in various domains, such as health care, labour environment, education, food production and restaurants, with the slogan "more effect, less burden". This sets targets for different inspectorates to both reduce the overall burden of inspection by 25% and to ensure a more effective and efficient approach (Inspection Council Bureau, 2009). In a similar vein, within the United Kingdom the Scottish Government commissioned a "reducing the burden of scrutiny action group" to examine possible ways to reduce the workload created by external evaluation efforts in the public sector (RBAG, 2008).[1]

Figure 6.1 Conceptual framework for school evaluation



Paradoxically, this lighter touch external school evaluation approach has further increased the importance of the school's self-evaluation activities and introduced new evaluation demands at the school level. There is, therefore, a concern at the national level to provide supports to these activities, typically via the provision of guidelines and tools for self-evaluation and via systems to feedback results from national student assessment activities to schools for use in their self-evaluation. The provision of benchmarked student performance data to schools is gradually gaining importance across countries. Further, the use of online data systems aims to increase the efficiency of compliance-related reporting and to reduce the time this takes for schools.

Finally, there is a shift in the focus of external school evaluation. Compliance is no longer the sole objective and there is an increasing focus on the quality of teaching and learning. This may be explicit via a judgement on the school's capacity to improve or implicit via the less frequent external supervision of schools judged to have good capacity for self-evaluation or review, thus representing a trust in the school's ability to improve.

Do external school evaluations lead to school improvement?

There is a lack of research into the impact of external school evaluation on school improvement. The only country with any research tradition in this area is the United Kingdom (Ehren and Visscher, 2008), although more recently research has started in Germany, Korea, the Netherlands and New Zealand. The Flemish Inspectorate of Education refrains from measuring its impact due to the difficulty of such research: external school evaluation has direct and indirect effects, as it fosters a school's awareness of its autonomy and accountability to improve its own quality (Flemish Ministry of Education and Training, 2010).

The purpose of external school evaluation and the nature of follow-up by external evaluators

Matthews and Sammons (2004, p. 153) in their in-depth evidence-based evaluation of the English inspectorate in the United Kingdom argue that assumptions that external school evaluation has a direct effect on school improvement are unrealistic without changing the nature of external school evaluation and giving external school evaluation bodies greater powers of follow-up or intervention. External school evaluation does not promote improvement by direct intervention, but rather by professional influence, fair and accurate reporting and informed analysis and comparison (idem). Indeed, Dedering and Müller (2010) argue that – in contrast to existing external school evaluation systems in England and the Netherlands – the purpose of recently introduced external school evaluation mechanisms in Germany is mainly for school improvement including an advisory and support function. They present research evidence from a survey administered to school principals in North Rhine-Westphalia that external school evaluations are discussed by a large group of stakeholders and are leading to the planning and implementation of school improvement and development actions.

External school evaluation impacts different schools in different ways

Existing research suggests that external school evaluation has differing impact on schools and that certain conditions are associated with schools accepting and acting on feedback from external school evaluation. For example, research on the impact of external school evaluation in England within the United Kingdom shows: this did lead to change in internal school structures in schools that had either received a negative

assessment or had areas to improve (Ouston et al., 1997; Kogan and Maden, 1999); secondary schools with lower or higher than average achievement did see slight improvement (Shaw et al., 2003); and the most and least effective schools made the most use of external school evaluation results, but that external school evaluation had made a substantial improvement to the education system as a whole (Matthews and Sammons, 2004). Parsons (2006) found that there was variability in how external reviews were conducted and received by schools in New Zealand, but judged the influence of external evaluation to be "pervasive, multi-faceted and subtle".

Clear feedback from external evaluators and acceptance of feedback by schools

There is evidence from different countries that the nature of feedback from external school evaluation has an important influence over its impact on school improvement. Matthews and Sammons (2004, p. 164) identify clear reporting of external school evaluation results and recommendations for improvement to be an important condition for the implementation of recommendations made by external evaluators. In ten case studies in Dutch primary schools, Ehren and Visscher (2008) found that all schools used external school evaluation feedback and six months after the external evaluation were still carrying out improvement plans and had already launched improvement initiatives that were relatively easy to implement. A combination of factors were identified as contributing to this: an assessment by external evaluators that certain points of the school's provision were "unsatisfactory", together with feedback from external evaluators on these weak points and agreement between the external evaluators and the school on improvement activities. Individually these factors did not explain the number of improvement initiatives launched by a school after an external evaluation. Therefore, the nature of feedback from external evaluators had a greater impact on school improvement than the amount of feedback they provided.

Two recent studies in the Netherlands also find that schools make use of external school evaluation reports and school quality report cards and that these do impact school policies and management (Bekkers et al., 2012; Janssens, 2012). In New Zealand, Nees (2007) studied six schools in the Wellington area following an external school review and found that all schools had made progress towards achieving recommendations made in external school review. In Korea, feedback from schools indicates that they find external evaluation reports considerably useful as they provide schools with practical assistance when designing education plans for the next year and establishing mid- to long-term school development plans. Schools also report appreciation of the dissemination of best-performing school cases (Kim et al., 2009). In Sweden, an audit of the external school evaluation process between 2003 and 2006 found that most schools constructively used feedback from external evaluation to improve their work and that the most important impact was that external school evaluation had brought about improvement sooner than would otherwise have been the case (Ekonomistyrningsverket, 2006).

Expectations that schools follow up on external evaluation results and school capacity to do so

There is also evidence that an external expectation for schools to follow up on feedback and school capacity to follow up on results play an important role. Matthews and Sammons (2004, p. 164) identify the following main conditions for the implementation of recommendations from external school evaluation: "understanding and acceptance of the findings by the provider; leadership that can generate and implement a

strategy for implementing inspection outcomes, including effective action planning; identification of any resources and support needed to effect improvement; planned external follow-up to assess the progress made; high stakes, where inspection has the potential to affect funding or public esteem for the provider." In the Flemish Community of Belgium, follow-up external evaluation of schools receiving extra funding under the equal opportunities policy found considerable improvement in schools in 2008 that had received negative evaluations from the Flemish Inspectorate of Education in 2005 (Flemish Ministry of Education and Training, 2010). The policy included requirements for schools to self-evaluate their use of funding and its impact; plus the recommendations from external school evaluators and subsequent follow-up had fostered schools towards more policy-oriented reflection and practices.

Ehren et al., 2013 find that although expectations set in external school evaluation and stakeholder sensitivity to its results are related to schools accepting feedback from external school evaluation, the feedback is not connected to school improvement actions, i.e. the results indicate that accepting external school evaluation feedback does not motivate schools to improve (see Figure 6.2). This "unsettling" finding may be due to a lack of communication from external school evaluation bodies on the expectation that feedback is used by schools for improvement, or a lack of capacity – or even resistance – to act on feedback at the school level (Ehren et al., 2013). In the French Community of Belgium, teacher representatives report that the acceptance of feedback from external school evaluation by pedagogical teams within schools is an important aspect for integrating feedback into the school's evaluation culture (Blondin and Giot, 2011).

In Korea, there is concentrated management and follow-up of schools that fail to reach certain quality levels (Kim et al., 2010). External school evaluation is conducted by metropolitan/provincial offices and some differentiate external school evaluation, i.e. schools gaining excellent results are exempted from external evaluation, but underperforming schools receive focused external evaluation and consulting. Linking external school evaluation results to requirements for external school evaluation and other administrative measures is found to be an effective means of raising the impact of external school evaluation (Jung et al., 2008). Almost all metropolitan/provincial offices of education use a system of rewards to provide schools with incentives according to their evaluation results. However, the absolute amount of incentives and the width of level-differentiation are not very significant and so the system is judged to have no major impact on schools (Jung et al., 2008). Further, feedback from school principals indicates that the incentive and reward system has little relevance to school improvement (Kim et al., 2009). In the Netherlands, the adoption of a risk-based approach to external school evaluation with a targeted focus on underperforming schools has seen an initial decline in the number of "very weak" schools within the system (Inspectie van het Onderwijs, 2010a, 2010b). In Ireland, a School Improvement Group was established in 2008 to follow up on underperforming schools and has found that this is a lengthy process, but that such targeted follow-up is starting to see results in schools (see Box 6.18).

Setting expectations of school quality, reporting findings to stakeholders and stimulating school self-evaluation

Of course, the stated purpose of external school evaluation often includes school improvement along with the major purpose of holding schools accountable by controlling aspects of their provision and quality. An ongoing research project funded by the European Union has analysed official documentation from external evaluation bodies (inspectorates) in six systems (Styria in Austria, the Czech Republic, Ireland, the

Netherlands, Sweden and England in the United Kingdom) and conducted interviews with representatives from external evaluation bodies and ministries to help map out how the design of each external evaluation system intends to impact school improvement (Ehren et al., 2013). Three common factors are identified as being expected levers for school improvement:

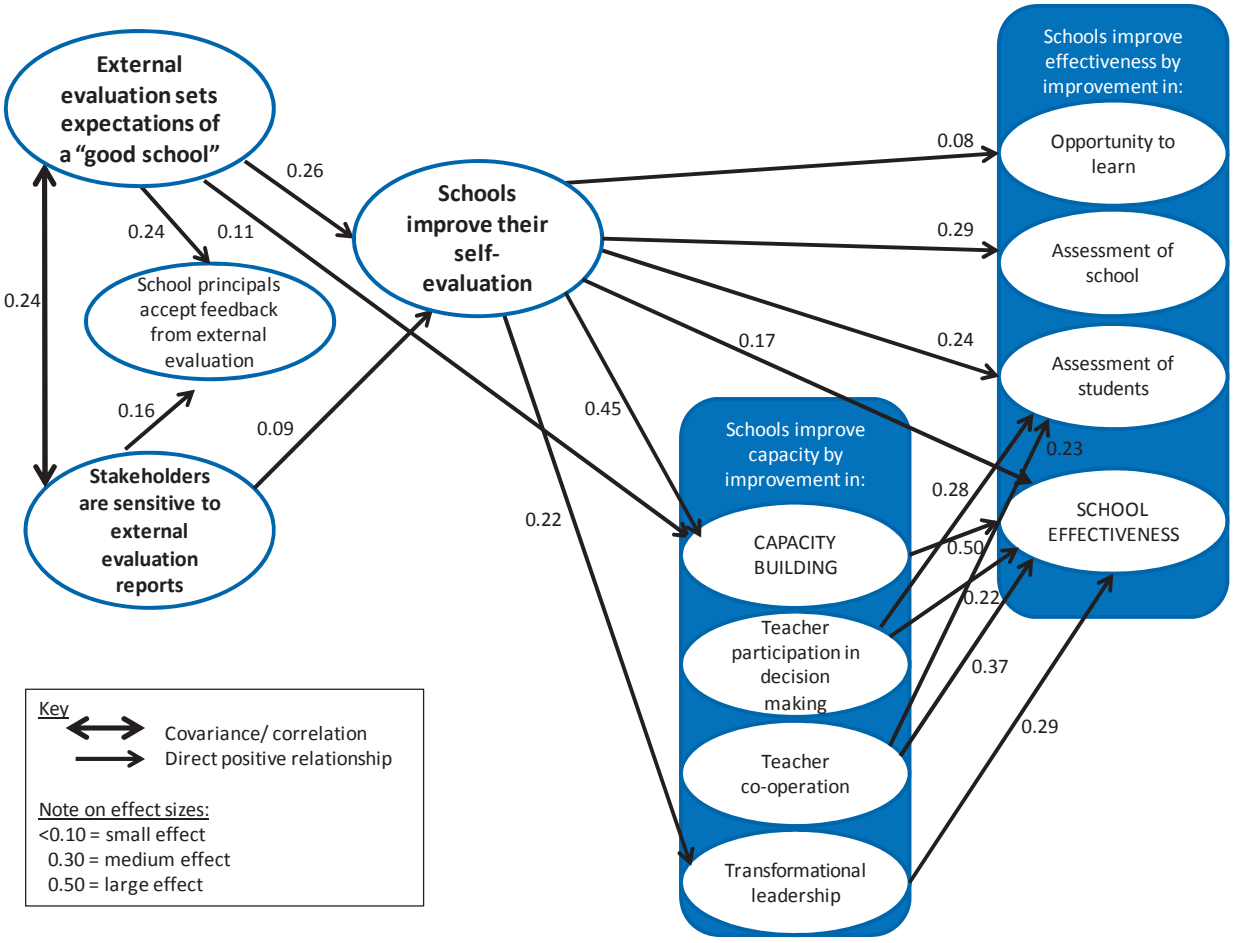
- External school evaluation sets expectations on school quality (i.e. with evaluation criteria and standards indicating a "good school");
- The results of external evaluation are shared with stakeholders (school boards/management, parents and students), stakeholders are sensitive to the results and this leads to pressure for improvement;
- External school evaluation promotes and stimulates improvement of school self-evaluation processes.

A preliminary testing of these theoretical levers has been conducted via the analysis of feedback from surveys administered to school principals in representative samples of schools in the six participating European systems (about 2 200 responses). Results indicate that clear expectations in external school evaluation and stakeholder sensitivity to the results of external school evaluation are correlated and are strong determinants of improvement actions reported at the school level (see Figure 6.2).[2] These findings suggest that where external school evaluation sets clear expectations, norms and standards and stakeholders are engaged with and knowledgeable about the external evaluation process, this has significant impact on schools (Ehren et al., 2013). Expectations set in external school evaluation and stakeholder sensitivity to the results of external school evaluation are also significantly related to schools improving their self-evaluation processes. There is also a direct influence of expectations set in external school evaluation over schools to build their capacity for improvement. These findings suggest that schools improving see systematic self-evaluation as a vital developmental strategy (Ehren et al., 2013).

Importantly, the findings indicate that various processes stressed by external school evaluation bodies to stimulate school improvement, such as school self-evaluation, transformational leadership and collaborative staff activities are important and effective (Ehren et al., 2013). Improvements in school self-evaluation are related to many school improvement actions. Figure 6.2 presents two distinct blocks of improvement actions at the school level. The first relates to the broad concept of improvements in capacity building, comprising improving teacher participation in decision making, improving teacher co-operation and improving transformational leadership. The second relates to the broad concept of improvements in school effectiveness, comprising improving opportunity to learn, improving assessment of the school and improving student assessment. School principals reporting that they are implementing or improving their self-evaluations, also report: taking more actions to build their capacity for improvement and change, notably, improving their transformational leadership; and improving their assessment of the school and students. In turn, schools reporting that they take more actions to build capacity for improvement and change, also report taking more actions to improve school effectiveness.

Figure 6.2 Evidence on reported improvement actions in schools following external school evaluation

Path analysis of school principal reports on external school evaluation in Styria in Austria, the Czech Republic, Ireland, the Netherlands, Sweden and England in the United Kingdom



Source: Ehren et al. (2013).

Research in New Zealand also indicates that the focus on school self-review during the external review process can effectively stimulate schools' approaches to improvement. Schagen and Wylie (2009) found that three-quarters of the primary school principals in their survey reported that the external review's affirmation of their approach to improvement was the main outcome of the most recent external review. According to feedback received from a survey in 2008-09 sent to a random sample of school principals 15 months after an external review, their perceptions of the review process and final report were that it had been useful in informing school developments and had most impact on improving the school self-review, building on strengths and addressing weaknesses.

Evaluation of the external school evaluation process in Korea has shown that it gives schools the opportunity to reflect on their educational activities and it promotes information sharing and exchange among school members, and helps them broaden their interest and understanding of overall school affairs (Jung et al., 2008). The majority of teachers report that school evaluation gives the opportunity to refresh the school atmosphere, and that it provides a venue for discussion and consultation on the school's strengths and weaknesses.

In Portugal, feedback collected by the Inspectorate from schools inspected in 2008/09, indicated that the majority found a positive impact on the development of their self-evaluation process (Portuguese Ministry of Education and Science, forthcoming). At the end of the first school inspection cycle (2006-11), school external evaluation was judged to have contributed to the development of self-evaluation in schools. In the French Community of Belgium, school principal representatives report that the focus of external evaluations on student results is a way to bring together the teaching staff to work on improvement (Blondin and Giot, 2011).

Unintended impact on schools

The above discussion examines evidence on the expected impact of external school evaluation. However, research also reveals examples of external evaluation impacting schools in unexpected ways. Unintended impact can include, for example, schools undertaking extensive preparation for external school evaluation and paying less attention to the teaching and learning process during that period (e.g. Rosenthal [2004] suggests this as a possible explanation of the finding that secondary schools had a drop in student achievement in the year of inspection) and undue stress for school staff in anticipation of an external school evaluation (e.g. Döbert et al. [2004] argue that this was the case for the assessment of school quality in France). An overview of empirical studies (mainly in the United Kingdom) finds that school staff report complaints of preparations for external school evaluation being stressful and time consuming, but finds no empirical studies on potential limitations on school diversity and innovation, via a "teaching to inspection" phenomenon (De Wolf and Janssens, 2007). Preliminary findings from an ongoing project funded by the European Union suggest that schools reporting they had accepted feedback from external school evaluation bodies did not perceive unintended consequences of external school evaluation (Ehren et al., 2013). However, schools reporting that they feel the external school evaluation body determines their expectations of good education, also reported unintended consequences of external school evaluation. Results indicate that new teaching approaches and curriculum experimentation may be hindered by school principals' concerns that these could distract staff from concentrating on meeting the expected external school evaluation standards (Ehren et al., 2013).

Governance

This section examines the purpose of school evaluation, which bodies are responsible for devising and organising school evaluation, the extent to which a legal framework is set to specify school evaluation activities, other policies systems use to stimulate school evaluation activities and how external school evaluation and school self-evaluation are articulated.

Purpose of school evaluation

As with all components of the evaluation and assessment framework, school evaluation has two major functions:

- **School development**: School evaluation identifies strengths and areas for school development with the aim to improve teaching and learning within schools, to close achievement gaps between schools and to enhance the performance of all students. This would necessitate a robust evaluation of the processes and strategies associated with student learning to allow the identification of areas for

school improvement. School evaluation can also be used to judge a school's capacity to enhance its effectiveness and therefore improve student learning.

- **School accountability**: School evaluation generates information for education authorities, the school community and other stakeholders to hold schools accountable for their performance. Information may include how a school complies with national standards and regulations, professional judgement on the quality of the services provided by the school, the learning outcomes and progressions of students within the school and judgements on the quality of outputs of a school based on its resources and other inputs ("added value" and "value for money").

Across countries, aspects of both purposes of school development and accountability can be found to varying degrees in self-evaluation and external school evaluation activities and in the comparison of schools on different performance measures. Typically, however, the last two are more associated with accountability purposes and school self-evaluation processes more with school development purposes.

School evaluation frameworks can draw on a significant body of research over the past 30 years that has defined the characteristics of effective schools and the processes to improve school effectiveness. This can help to devise an overall approach to school evaluation that strives for school improvement by use of evaluation for both school accountability and school development. Yeung (2011) provides a helpful summary to compare the school effectiveness and school improvement literature in the context of school evaluation (Table 6.1).

Table 6.1 Purpose of school evaluation

Accountability purpose "School effectiveness"	Development purpose "School improvement"
Focus on schools	Focus on teachers
Focus on school organisation	Focus on school processes
Data-driven, with the emphasis on outcomes	Empirical evaluation of effects of changes
Quantitative in orientation	Qualitative in orientation
Lack of knowledge about how to implement change strategies	Exclusively concerned with change in schools
More concerned with change in student outcomes	More concerned with journey of school improvement than its destination
More concerned with schools at one point in time	More concerned with schools as changing
Based on research knowledge	Focused on practitioner knowledge
Concerned with schools that are effective	Concerned with how schools become effective
Static orientation (school as it is)	Dynamic orientation (school as it has been, or might be)

Source: Reynolds et al. (1996) and Chapman (2005) in Yeung (2011).

The mix of school evaluation policies needs to fit into the wider governance context in a school system. For example, the Swedish system is strengthening external national control at a time of market orientation and greater choice and privatisation in the Swedish school system (Ronnberg, 2011). To better fit a wider move to promote "good governance" in the Netherlands, the results of external school evaluations are now communicated directly to the competent school authorities and not directly to schools (Scheerens, et al. 2012). In France, there has not been much support for the development

of whole-school evaluation given the limited autonomy for primary schools in particular, but also for secondary education providers (Dos Santos and Rakocevic, 2012). However, central education authorities (general inspectorate and directorates for evaluation and school education) are committed to working towards a better school external evaluation system. They put special emphasis on close co-operation between schools and regional education authorities. Boxes 6.1 and 6.2 present some examples of country initiatives to strengthen the purpose for school development or for school accountability.

Box 6.1 Strengthening the focus on school development

Korea: Shifting the purpose from compliancy to providing direction for improvement

As a result of strengthened school autonomy, Korea introduced a school evaluation system in 1996 to enhance the quality of education and hold schools accountable (Kim et al., 2010). The focus was on school compliancy with national and local policies. From 2000-2004 there was a dual system of national "qualitative reviews" on a sample of schools and Metropolitan/provincial offices of education (MPOEs) evaluated how well schools implemented local policy projects. In 2005 a national model for school evaluation was introduced comprising an element of "external review", a national school evaluation framework with core common quality indicators and both external and self-evaluation activities. However, external school evaluation is conducted solely by the MPOEs. MPOEs include local indicators in addition to the common national indicators. The national school evaluation framework was introduced to allow the evaluation of the whole school system against the standard of the ideal school education and to provide direction for school education reform. Prior to this, supervision and inspection focused only on checking whether schools were carrying out policy projects specified by central and local education offices and abiding by educational laws.

Luxembourg: Stimulating school use of data for development planning

Since 2009, fundamental schools (ISCED 0 and 1) have been required to set a 4-year development plan. This is within the context of a reform to focus fundamental education on competency development at different key stages and has been accompanied by the introduction of national student assessments at two of the four key stages (Cycles 3 and 4). The Ministry has strongly followed schools in their development planning and by the end of 2011 all fundamental schools had developed a 4-year plan (Shewbridge et al., 2012). A major focus from the Ministry has been to help schools with analysing data and it offers feedback from national assessments, other assessment tools, as well as advice and analytical expertise. Each year the school team should evaluate its implementation of the school development plan. This implies reviewing the achievement of annual school objectives and adapting those to be implemented in the following year.

Denmark and the Flemish Community of Belgium: Introducing robust performance data for schools to use in self-evaluation

In 2010, Denmark introduced a suite of computer-based national assessments with the aim of providing schools and teachers with rapid feedback on how students perform in discrete areas of the national common objectives. Such information is fed back to schools via analytical software which can be used to compare student performance to national averages and various student groupings within the school. Both the speed of feedback of results (the day after the student is tested) and the flexibility of analytical functions in the software used heighten the relevance of student performance data to school self-evaluation activities (Shewbridge et al., 2011b).

The Flemish Community of Belgium provides feedback reports to schools on their performance in both the national sample assessments and international assessments. Further, it offers schools which have not participated in the national sample the possibility to conduct the assessments and to receive feedback. The results included in the feedback reports are benchmarked against averages in the Flemish Community, as well as adjusted to allow comparison to schools with similar student characteristics (Shewbridge et al., 2011a).

Box 6.2 Strengthening the focus on school accountability

Sweden: Strengthened national capacity to conduct external school evaluation

External school evaluation was introduced at the national level in 2003 and initially conducted by the National Agency for Education. In 2008, a new agency was established: the Swedish Schools Inspectorate. The first round of external school evaluation was conducted over a 6-year period. The Swedish Schools Inspectorate plans to evaluate each school on a 3-year cycle. External school evaluation follows nationally established standards and external school evaluation reports are published on line to complement performance information in the national reporting systems (Swedish Ministry of Education and Research, 2010).

Netherlands: School compliancy re-emphasised in external school evaluation

A revision to the Supervision Act in 2012 corroborates the responsibilities specified in the 2010 "Good governance, good education" Act, i.e. that the competent school authorities are held accountable for student results and school governance and financial compliance (Scheerens et al., 2012). External school evaluation criteria (as specified in the "school inspection framework") have always included school compliance to educational laws, but now explicitly state certain laws, e.g. parental participation in school decision making. Further, the 2012 Act introduces the new aspect of the evaluation of a school's teacher personnel policy.

Australia: Easy access public website providing performance data on schools

A major school reporting system was launched in Australia in 2010 (*www.myschool.edu.au*). The My School website continues to evolve and each annual release includes more information on schools. The major impetus initially was to provide transparent performance information on each school in Australia. My School currently presents a suite of information on school context and mission, in-depth presentation of the school's performance on the National Assessment Program – Literacy and Numeracy (NAPLAN), as well as funding information for the school. There is also a mechanism for visitors to the site to compare both a given school's performance and its performance progress to the national average and to a group of statistically similar schools (see Box 6.20).

School accountability takes multiple forms

The OECD Review has revealed a complex mix of different types of school accountability co-existing in many countries. Schools are increasingly held accountable to multiple levels of educational authorities and to the wider community in which they are involved (Faubert, 2009). In most countries, schools provide information to the relevant public authorities (local or national) (OECD, 2011). This is the traditional vertical or hierarchical external accountability, but schools may also be accountable to their supervisory boards (OECD, 2011). The publication of school performance measures and the results of external school evaluation aims to introduce another form of external accountability, in which schools are accountable to the market, i.e. parents choosing a school. More recently, a horizontal accountability has emerged and schools provide their community and stakeholders with insight into their processes, choices and results (Faubert, 2009). This is also linked with the development of a strong role for school self-evaluation. For example, in Scotland in the United Kingdom, external school evaluators validate a school's self-evaluation results and only where there are concerns about the school's effectiveness in serving its community will there be more intensive external school evaluations. This reflects the horizontal accountability of the school to its students, their parents and the community (Hutchinson and Young, 2011).

Across the OECD, school compliancy with legislation is an important part of external school evaluation. This may be complemented by an evaluation of the quality of the processes developed by schools in order to meet the goals set by educational authorities.

Where evaluation exclusively monitors compliance with regulations, this can include aspects such as the composition of classes, the tasks and workload of teachers, the use of school infrastructure, and the management of human and budgetary resources, but it does not include educational tasks (Faubert, 2009).

In addition, most countries include a focus on outcomes, including student cognitive and social skills (Faubert, 2009). The major information base for outcomes at the school level is aggregate performance measures on student results in national assessments or examinations. In the United States, this is the major focus in the approach to school accountability. Among the OECD Review countries, Hungary and Mexico have a strong reliance on school performance measures.

However, in most countries there is a mix of a focus on processes and outcomes. Masters (2012) reviews research on outcomes-based incentive schemes and argues that a system to reward school improvement should be based on both outcomes-based measures and practice-based measures. It is logical to use evidence of improved practices and processes in a system that aims to improve school quality.

Risks that compliancy dominates school evaluation

There is a risk that external evaluation may be predominantly associated with compliance to procedural requirements, instead of with school improvement. This may be inherent in both the design of the evaluation system and what is assessed and choices to introduce shorter inspections drawing heavily on school documentation. There is a risk that external school evaluation does not yet place adequate focus on teaching and learning and misses the opportunity to contribute to school improvement. This means that the external school evaluation process sends ambiguous signals about what matters and forms judgements on proxy indicators, the evidence for which is open to manipulation and misrepresentation.

Increased pressure stemming from external school evaluation might create incentives for mere compliance to administrative requirements at the expense of improvement and innovation (Faubert, 2009). An approach that is largely top-down and that imposes changes on schools is believed to create a "culture of compliance" among teachers at the expense of innovation (Datnow et al., 2000; Leithwood et al., 2000). It may also be the perception of school principals that external school evaluation is only an exercise in compliancy to ensure continued recognition or funding and that it is not a useful stimulus for school improvement. Such perceptions may reflect a lack of follow-up by external authorities on the implementation of recommendations within the evaluation reports. The association of external school evaluation with compliancy may also mean that school principals do not promote and seriously discuss the results of external school evaluation with the full school staff and parents, thereby severely reducing their impact on school improvement.

The purpose of the evaluation exercise and who has control over it are sensitive issues (Simons, 1987; Ball, 2003). Looney (2011) finds that the misalignment of views among different stakeholders on the role and purpose of evaluation poses significant challenges to the effective use of evaluation for reform. A study commissioned by the National Union of Teachers in the United Kingdom (MacBeath et al., 1996, p. 92) stated four key priorities for improving the school evaluation process: self-evaluation should be central in any national approach to school improvement; accountability and self-improvement should be seen as two strands of the one inter-related strategy; provision of time and resources have to feature as a key issue in school improvement; and external school

evaluation should continue to be a feature of the drive towards school improvement, but as part of a collaborative strategy with schools and local authorities.

West, Mattei and Roberts (2011) argue in the presence of multiple types of accountability for schools in England within the United Kingdom, that hierarchical and market accountabilities (e.g. schools demonstrating improvements on measures published in school league tables), as well as legal accountability (e.g. schools facing potential sanctions) carry most influence. As such it is argued that compliancy dominates the school accountability system. Network, participative and professional accountability are perceived to be comparatively weak forms of accountability, although they could help to foster greater social cohesion.

Within the United Kingdom, Scotland has a clear policy for school self-evaluation to be central to school self-evaluation. However, a change of culture requires sustained effort and capacity building. Croxford et al. (2009) comment that a strengthened role for self-evaluation has been hindered by: the historical context of strong central influence; and other policies regarding the setting of targets by local authorities and the use of standard performance measures. They argue that school self-evaluation has become an accepted procedure in schools, but that some undertake this "enthusiastically", while others "treat this with cynical compliance". Hutchinson and Young (2011) identify that the new policy emphasis on assessment for learning (encouraging teachers to make professional judgements on student learning progress and to minimise reliance on standardised tests) articulates well with the external school evaluation policy. However, it is necessary to build professional and public understanding of the new assessment and evaluation models and to build capacity for using and interpreting data. Teachers will need appropriate professional development and support from authorities to avoid the risk of slipping back into the old paradigm of using external tests to provide robust, reliable and objective evidence to external school evaluators and authorities (University of Glasgow, 2008). This echoes warnings that reliance on a heavy test-based accountability system may threaten professional development and capacity building and result in a culture of dependence (Earl et al., 2003).

In Norway, the Norwegian Labour Inspectorate conducts inspections of the health, environment and safety in Norwegian schools, but it does not focus on the education provided in schools or the quality of teaching and learning. Typically, local and regional authorities, as part of their obligation to monitor schools, take note of school results, sometimes require schools to submit annual strategic plans and/or improvement plans and occasionally visit schools to interview senior staff and check compliance with legislation. They do not generally undertake more in-depth school reviews or inspections involving the direct observation and evaluation of the quality of teaching and learning (Nusche et al., 2011a). Similarly, the typical approach of the regional representatives of the central authorities in monitoring local approaches is to simply monitor results and check compliance at an administrative level.

In Denmark, the annual requirement for local authorities to produce a quality report has been beneficial in promoting dialogue between local authorities, school boards and schools and making the work of schools more transparent (Shewbridge et al., 2011b). The use of a national template, but with local adaptations has promoted a sense of ownership of the quality reports, providing that schools and school boards actively participate in their development and see value in the data and information they contain. The evolutionary process of developing these reports involves dialogue about what is worth reporting and how it can be measured and estimated. However, the reports typically contain indicators

that are not sufficiently scientific or focused to diagnose the strengths and weaknesses of the core processes relating to teaching and learning and leadership. This reporting process does not guarantee an evaluation of the quality of teaching and learning and notably does not necessarily stimulate a culture of teacher appraisal and classroom observation.

In the Slovak Republic, there have been increased demands on the State Schools Inspectorate to conduct different types of external school evaluations, including "information inspections" (Shewbridge et al., forthcoming). These types of external school evaluation focus on documentation, are conducted in a single day and do not include classroom visits. For example, recent information inspections have included verification of the content of the school educational programme against the national educational programme. Schools and other stakeholders report experiencing information inspections as bureaucratic exercises placing demands on school time, but not providing useful feedback for the school. Although schools are supposed to have an external whole-school evaluation (complex inspection) every five years, this external evaluation cycle appears to be under pressure and some schools may wait longer. During an external whole-school evaluation, external evaluators conduct a number of classroom observations using a stable analytical observation and judgement instrument and often ask the teacher to give a brief self-evaluation of the lesson before giving feedback. Stakeholders during the OECD review in the Slovak Republic expressed strong support for these whole-school evaluations.

In Mexico, there is no systematic external school evaluation. However, there is a long-established tradition of oversight of school work by supervisors and other personnel external to the school, but their role has been largely associated with ensuring schools' compliance with regulations and other administrative tasks. The traditional role and functions of supervisors, relating to regulations, control, administrative operation of schools and supervision of the political and ideological standpoints of teachers do not appear to have evolved or to respond appropriately to the needs of the education system (Santiago et al., 2012). However the OECD Review did reveal some examples of supervisors providing feedback to school principals and teachers and their role could evolve to one that can support school self-evaluation.

Responsibilities for external school evaluation

Responsibilities for devising external school evaluation lie firmly in the realm of educational authorities (central, regional/provincial and local) and specific external school evaluation bodies such as inspectorates or school review bodies. In the majority of education systems operating a system for external school evaluation, this is devised and organised at the central or state level (Table 6.2).

In education systems operating a system for external school evaluation, the major capacity lies with:

- **Central or state authorities**: The Ministry of Education is responsible for external school evaluations in Denmark, Iceland, Israel and Norway. In Australia, arrangements vary among states and territories, but are generally managed by government departments. For example, in Western Australia there is an Expert Review Group within the Department of Education and Training, but in Victoria external reviews are commissioned to external, accredited individuals but review processes are evaluated by officials in the Victorian Department of Education and Early Childhood Development (Santiago et al., 2011). In Ireland and Northern Ireland in the United Kingdom an inspectorate sits within the Department of Education.

- **A specific body beyond the Ministry of Education**: Typically, in European systems this is a national or state level school inspectorate *(*e.g. the Slovak State Schools Inspectorate, the Flemish Inspectorate of Education, etc.). In New Zealand this is the national Education Review Office. Sweden established the Swedish Schools Inspectorate in 2008, following the commencement of external school evaluation activities in 2003 by the National Agency of Education. In Germany, all 16 federal states have implemented some form of external school evaluation between 2004 and 2008. In six states there are specific institutions responsible for external school evaluations (Agency for Quality in Bavaria; Institute for School Development in Bremen; Department of the Institute for Educational Monitoring in Hamburg; Institute for Quality Development in Hesse; Agency for Quality in Rhineland-Palatinate; Department in the Institute of Education in Saxony) (Dedering and Müller, 2010).

In Chile, the Czech Republic and Poland, external school evaluation responsibilities are divided among the central and provincial/regional levels. Korea provides an example where such division of responsibilities was formerly the case (2000-04), but responsibilities were changed to reduce overlap of evaluation activities (Kim et al., 2010). Metropolitan/provincial offices of education now have full responsibility for conducting external school evaluations, but these are based on a national school evaluation framework. Other countries where responsibilities lie at the provincial/regional level include Austria and France (secondary schools) and Turkey (primary schools). In France, where there is not a unique school evaluation protocol, the school evaluation bodies based in the regions participate in school evaluation, in addition to different directorates of the Ministry of National Education and to the national inspectorate. The regional inspectors take stock of the situation in the school with the school principal, usually during their visit for teacher appraisal. The ministerial directorates (especially the Directorate of Evaluation, Forecasting and Performance, DEPP) are responsible for elaborating performance indicators. The national inspectors evaluate the implementation of particular reforms and policies in different types of schools, for example an evaluation in 2012 of educational reintegration programme schools.

It is far less typical for external school evaluation responsibilities to lie at the local level. In Austria and France, this is the case for the external evaluation of primary schools. In France this is the traditional individual inspection by local inspectors called National Education Inspectors (IEN), but changes are currently being discussed. In Norway, external school evaluation responsibilities lie at the local level for primary and lower secondary education. However, since 2009 regional representatives of the central authorities are obliged to monitor local approaches to school evaluation. It is of note that in Luxembourg there is no external school evaluation, but there is a system of "inspectors" at the primary level. These are the hierarchical heads of primary schools, but they are not physically located at the school and are responsible for all primary schools in a given local authority.

In both Finland and Hungary, national school inspectorates were abolished (in 1991 and 1985, respectively) and no systematic external school evaluation is conducted, although in both cases there is nationally recognised capacity for external school evaluation, should schools choose to commission this. It is of note that in Hungary governance structures are changing from 2013 and that this will have implications for school evaluation (see Chapter 2). In Denmark, it is possible for private schools to nominate a person of their choice to conduct external school evaluation, but this person must be accredited by the Danish Ministry of Education.

Table 6.2 Responsibilities for external school evaluation

	Major bodies responsible for conducting external school evaluation
Australia	Some states organise external review of public schools typically conducted by special groups within the government department of education. May also conduct external reviews of private schools as part of the registration process. Evaluations may also be organised at the local level by some non-government school systems (e.g. Catholic sector).
Austria	Local (ISCED 1); local and provincial/regional (ISCED 2); regional (ISCED 3)
Belgium (Fl.)	State: Flemish Inspectorate of Education
Belgium (Fr.)	State: General Inspection Services (*Service général de l'inspection*)
Canada	The Atlantic provinces require the external review of schools which is typically conducted by a member of the department of education and representatives from other school boards.
Czech Republic	Central: Czech School Inspectorate and regional (ISCED 2 and 3) and municipal (ISCED 1 and 2) education authorities
Denmark	Ministry of Education monitors school performance in lower and upper secondary schools and in all private independent schools. As of 2010, private schools can either nominate an accredited external inspector or conduct school self-evaluation.
Estonia	Central
Finland	No central external school evaluation (School Inspectorate abolished in 1991). Education providers may commission and pay for external school evaluations by the Education Evaluation Council (the same body that conducts national evaluations).
France	Local (ISCED 1): school inspection visits undertaken by National Education Inspectors (IEN) Regional (ISCED 2 and 3): Regional Pedagogical Inspectors (IAIPR); General Inspectorate of National Education (IGEN); General Inspectorate of the Administration of National Education and Research (IGAENR)
Germany	State: in nine states the Ministry of Education or other government department; in six states an institute external to the Ministry of Education; in one state an autonomous body of school governors.
Greece	No external school evaluation
Hungary	No external school evaluation (inspection activities abolished in 1985 Education Act). National Register of Experts contains names of individuals who are competent and permitted to perform external school evaluation.
Iceland	Central and local (ISCED 1 and 2); central (ISCED 3) – Ministry of Education, Science and Culture
Ireland	Central: Inspectorate of the Department of Education and Skills
Israel	Central: Ministry of Education
Italy	No external school evaluation
Japan	No external school evaluation (ISCED 1, 2 and 3 general.); Central (ISCED 3 [pre-]vocational)
Korea	Provincial: metropolitan/provincial offices of education
Luxembourg	No external school evaluation
Mexico	No external school evaluation
Netherlands	Central: Dutch Inspectorate of Education
New Zealand	Central: Education Review Office
Norway	Local (ISCED 1 and 2): municipalities. Regional (ISCED 3): counties. Local and regional authority representatives may visit schools to interview leadership, but generally do not conduct reviews or inspections. Regional representatives of central government (County Governors) ensure that local and regional authorities have an effective quality system in place.
Poland	Central and regional: Ministry of Education and regional superintendants
Portugal	Central: General Inspectorate of Education and Science (IGEC) established in current form in 1979
Slovak Republic	Central: State Schools Inspectorate
Slovenia	Central: The Inspectorate of the Republic of Slovenia for Education and Sport
Spain	State
Sweden	Central: Swedish Schools Inspectorate (2008); National Agency for Education (2003-07)
Turkey	Provincial/ regional (ISCED 1); Central (ISCED 3)
United Kingdom (England)	Central: Office for Standards in Education, Children's Services and Skills (Ofsted) established in 1992 as a non-ministerial government department
United Kingdom (Northern Ireland)	State: Education and Training Inspectorate of the Department of Education
United Kingdom (Scotland)	State: Education Scotland – Her Majesty's inspectors (2011); 2001-10 executive agency Her Majesty's Inspectorate of Education (HMIE)
United States	State, local and school

Source: OECD Review; OECD (2011).

Setting requirements for external school evaluation

The vast majority of OECD countries have established a legal basis for undertaking external school evaluation, although the extent and type of requirements set vary enormously. The only OECD countries in which there are no external school evaluation requirements are Finland, Greece, Hungary, Italy, Japan (with the exception of upper secondary vocational education programmes), Luxembourg and Mexico.

When setting a legal framework for external school evaluation, the typical approach across OECD countries is to set conditions for highly structured external school evaluations (Table 6.3). That is, specifying that similar evaluation activities should be completed at each school based on a specific set of data collection tools (OECD, 2011). Austria is the only system where external school evaluations are entirely unstructured, that is evaluation activities may vary depending on the strengths and weaknesses in different schools. In other systems, external school evaluations may fall between these two extremes in different ways.

Some systems establish requirements for annual reporting systems and these may to varying degrees include specifications on common indicators to be included. For example, in Norway a revision to the Education Act in 2009 established the basis for a degree of external evaluation. This aimed to address concerns that local authorities (ISCED 1 and 2) and regional authorities (ISCED 3) were not implementing adequate external school evaluation procedures (Nusche et al., 2011a). Local and regional authorities are obliged to develop a "quality framework" to guide their schools' self-evaluation activities and produce an annual status report. In addition, regional representatives of the central government are obliged to hold local and regional authorities accountable with regard to their duty to have effective quality monitoring systems in their schools. There is a centrally specified set of quality indicators (some mandatory and some recommended) that local and regional authorities are advised to include in their quality frameworks (and monitor in their schools), but local and regional authorities are also free to specify and add different quality indicators that are tailored to the local context (Norwegian Directorate for Education and Training, 2011). Further, the annual status report should form the basis of a programme of quality improvement and development activities for the year ahead. This is similar to a policy introduced by Denmark in 2006 for local authorities to produce an annual quality report on their schools. Again, the reports should address nationally specified indicators, but there are ongoing experiments to reduce the level of mandatory indicators and allow greater local flexibility. This is an attempt to respond to criticisms from both local authorities and schools of the burden that the quality report design placed on them (Shewbridge et al., 2011b). Indeed, the long established requirements for local authorities in Sweden to produce annual quality reports on their schools was dropped in 2010 due to similar concerns raised by local authorities on the amount of resources required to produce such reports (Swedish Ministry of Education and Research, 2010). However, the context in Sweden was different given a more highly developed mechanism for external school evaluation with the establishment of the Swedish Schools Inspectorate in 2008. Similar concerns are raised by educators in Korea, claiming that the burden of preparing a self-evaluation report distracts teachers from their class instruction hours and that the reported information is not usually related to the core educational activities of teachers (Kim et al., 2010). As of 2011, there is a new initiative in Korea to give more local freedom over the content of self-evaluation reports.

Table 6.3 Legal frameworks for external school evaluation

	Degree to which external school evaluation is structured and legal framework
Australia	Partially structured: all states and territories (for public schools) and non-government schools in the Catholic education sector provide school evaluation frameworks, although external evaluation for Catholic schools follows internal processes of school evaluation
Austria	Unstructured
Belgium (Fl.)	Highly/partially structured: 2009 Decree on Quality of Education ("differentiated" inspection in terms of intensity and frequency depending on the school quality) specifies that the focus of the inspection may change depending on the specific school, although all inspections are based on highly structured framework; whole-school evaluation since 1991
Belgium (Fr.)	Partially structured: 2007 revision of Inspection Act; Inspection of different study disciplines (e.g. mathematics); schools should be inspected in one discipline at least once every three years
Canada	Partially structured: the Atlantic provinces legislate School Improving Planning, which includes a final stage of external review (typically after 3-4 years) to validate a school's self-review against its school development plan.
Chile	Highly structured: Provincial Departments of Education (DEPROVs) responsible for the technical and pedagogical support, and the administrative and financial situation of schools that are under their jurisdiction (section 16 of Law No. 18,956)
Czech Republic	Highly structured: Education Act 2004 authorises the Czech School Inspectorate to inspect schools; 2008/09 inspectorate monitors school education programmes against national framework for education programmes. 2012/13 introduces a 4 year inspection cycle for schools at ISCED 1 and 2 (previously on a 3-year cycle). ISCED 3 schools remain on a 3-year cycle.
Denmark	Partially structured: system of local authority (municipal) quality reporting requirements introduced in 2006, based on common national indicators. Local authorities must publish an annual quality report on their schools and present an action plan for any school with identified quality concerns.
Estonia	Highly structured
Finland	No central external school evaluation
France	Partially structured
Germany	Highly structured: school inspections were implemented between 2004 and 2008 in all of the 16 federal states in Germany. Each state uses its own school quality framework including standardised criteria to evaluate "good instruction" and "good schools".
Greece	No external school evaluation
Hungary	Unstructured: No external school evaluation requirements, but nationally accredited evaluators available.
Iceland	Partially structured: Compulsory School Act (2008) strengthens evaluation and monitoring of school operations, specifies local authority supervision responsibility and requires 3-year plan for national evaluations.
Ireland	Highly structured: Education Act 1998, plus regulation in 2006 with respect to publication of inspection reports.
Israel	Highly structured
Italy	No external school evaluation
Japan	Highly structured (ISCED 3 pre-voc/voc.); no external school evaluations at other levels.
Korea	Highly structured: Primary and Secondary Education Act and the Enforcement Decree (1996).
Luxembourg	No external school evaluation
Mexico	No external school evaluation
Netherlands	Partially structured: Supervision Act (2002, revised in 2012) describes the inspection framework, interventions for failing schools and quality standards. Inspection of regulations specified in the Acts on Primary Education, Secondary Education and "Good education, good governance".
New Zealand	Partially structured: since 2009 differentiated approach to review schools with strong performance and self-review capacity every 4-5 years, schools performing well every 3 years and schools experiencing difficulty an ongoing review over a 1-2 year period.
Norway	Partially structured: local (ISCED 1 and 2) and regional (ISCED 3) authorities must establish and maintain a quality framework and prepare an annual status report including some mandatory indicators (Education Act revision in 2009).
Poland	Highly structured
Portugal	Highly structured: 2002 Evaluation Law established external and internal school evaluation requirements. First full inspection cycle completed between 2006 and 2011.
Slovak Republic	Highly structured: Act in 2003 (Inspectorate's role) and Decree on School Inspection in 2005 (types of inspection); the School Act 2008

Table 6.3 Legal frameworks for external school evaluation (*continued*)

	Degree to which external school evaluation is structured and legal framework
Slovenia	School Inspection Act (1996; last revision 2005) regulates matters of compliance with school regulations in pre-tertiary education
Spain	Partially structured
Sweden	Highly structured
Turkey	Highly structured
UK (England)	Highly structured: Education Act 1992 establishing Ofsted; Education Act 2005
UK (Northern Ireland)	Highly structured: The Education and Libraries (Northern Ireland) Order 1986 (Articles102 and 102A)
UK (Scotland)	Highly structured: starting point is external review of school's own self-evaluation conducted following a centrally devised framework
United States	Highly structured: the No Child Left Behind Act (2001) requires the public reporting of school performance on national standardised tests and specifies requirements for adequate yearly progress on these measures.

Note: Highly structured external school evaluation means that similar activities are completed at each school based on a specific set of data collection tools; unstructured external school evaluation comprises activities that vary at each site depending on the strengths and weaknesses of the school.

Source: OECD Review; OECD (2011).

In the case of the French Community of Belgium, schools may be inspected in different study discipline areas, although following a standard evaluation approach (Blondin and Giot, 2011). In the Flemish Community of Belgium, the inspection framework is based on common criteria, but schools may be evaluated against a subset of these criteria according to what is judged most pertinent to that school in terms of their quality improvement needs (Flemish Ministry of Education and Training, 2010). In both the Netherlands and New Zealand a common external school evaluation framework is used, but external school evaluations are differentiated depending on the evaluation of the school's quality. Such legal frameworks aim to better target external school evaluation to both the schools that are in most need of improvement and to particular areas of development required within each school. It is hoped that this will increase the impact of external school evaluations.

Balancing a need for regular external school evaluation and minimising the demands on school time

External school evaluation can be carried out at regular intervals (e.g. every four years), be a one-off event such as when the risk of underperformance is considered high, or may happen as the result of a complaint. However, systems operating external school evaluations often follow a set cycle and may specify the maximum period (usually number of years) between external evaluations conducted in a given school. For example, schools should have an external evaluation every five years in Slovenia and the Slovak Republic, every four years in the Netherlands and every three years in the Czech Republic. In the Flemish Community of Belgium, new regulations in 2009 saw the reduction in frequency of external school evaluations from at least once every six years to every ten years, although schools with identified quality concerns would be supervised more frequently (Shewbridge et al., 2011a). Ireland is an example of where there has been concerted effort to increase the frequency of external school evaluations. Reports from lower secondary teachers in the international TALIS survey indicated that just 43% were in schools where an external evaluation had happened over the past five years. In

response to this finding, the frequency of external evaluation has increased since 2009, "partly through the introduction of a range of new models of inspection and unannounced short inspections. In 2011, for example, inspections were conducted in over one-sixth of all primary schools and in over 600 of the 740 secondary schools in the country" (Irish Department for Education and Skills, 2012).

There may be a policy to introduce a "proportionate" approach to external school evaluation. This may still keep a regular cycle of external school evaluation in all schools, but include shorter school visits with smaller external evaluation teams in schools where there are no quality concerns detected. Or this may involve visiting better performing schools less often and schools with quality concerns more often. Moves to "proportionate" approaches to external school evaluation have taken place in the Flemish Community of Belgium, the Netherlands, New Zealand and in Scotland within the United Kingdom. Since 2009, the Flemish Inspectorate follows a "differentiated" approach to external school evaluation. This means that external school evaluations are differentiated according to both their frequency and their focus and coverage according to which areas of the external school framework are most pertinent to the particular school being reviewed (Shewbridge et al., 2011a). In Australia, Western Australia, New South Wales and Victoria have introduced variants of a risk-based determination for external reviews in public (government) schools (Santiago et al., 2011). Western Australia varies the length of the external review cycle depending on the school's performance, as annual reviews were deemed unnecessary for high performing schools. Victoria uses four increasingly intensive forms of external review depending on an assessment of a school's performance levels: negotiated, continuous improvement, diagnostic, and extended diagnostic. In New Zealand the frequency of external school evaluation is linked to an evaluation of the school's capacity to conduct self-evaluation (see Box 6.4).

External school evaluations may place significant demands on school time and resources. The time that a whole-school evaluation takes may not be worth its potential impact on school improvement. Some schools may benefit more than others from external school evaluations and may need more frequent evaluation than that scheduled in a regular cycle of external school evaluation.

In Hong Kong-China curriculum leaders, that is senior teachers within schools, report that they used certain coping strategies to deal with external school review demands, including groups to study external review criteria, planning and acting to meet performance indicators standards, preparing teachers to cope with classroom observation, help with compiling documentation and learning how to present evidence well (Yeung, 2012). However, the demands for external school review had also stimulated peer observation of teaching and engaging external consultants.

In the Netherlands, evaluations conducted by the Inspectorate of Education showed that the new approach of "risk-based inspection" entailed a lower administrative burden for schools, as well as providing efficiency gains for the Dutch Inspectorate of Education (Inspectie van het Onderwijs, 2010a, 2010b).

Setting requirements for school self-evaluation

In many countries (e.g. Australia, Germany and England within the United Kingdom) school self-evaluation activities have been initiated by individual schools or groups of schools through partnership with a university or school district (MacBeath, 2008). In Canada, there has been a distinct approach in developing evaluation practices from the bottom-up (Fournier and Mildon, forthcoming). For example, in the Atlantic Provinces in

Canada, school self-evaluation activities pre-dated legislation for school improvement planning. However, across countries there has been an almost universal focus at the national policy level to stimulate school self-evaluation. The vast majority of OECD countries have legal requirements in place for schools to conduct self-evaluation, although these vary significantly in nature (Table 6.4). For European systems, there has been a supra-national influence over the development of school self-evaluation. The European Parliament and Council of the European Union (2001) made a clear recommendation in 2001 for European countries to "encourage school self-evaluation as a method of creating learning and improving schools". In fact, nearly all European Union members already had some form of support for school self-evaluation in place at that time (Eurydice, 2004).

Table 6.4 Requirements for school self-evaluation

	Level responsible	Legal framework
Australia	State/territory and school	Partially structured: all schools must publish an annual report including school performance information on key outcomes and satisfaction, as well as contextual information about the school.
Austria	Schools and regional authorities	The 2011 Schools Inspectorate Act provides for the introduction of mandatory self-evaluation. School development plans are being piloted in some schools and will be mandatory for schools offering general education in 2013/14.
Belgium (Fl.)	No explicit self-evaluation requirements	2009 Decree on quality of education states that schools "need to be able to account for their efforts to monitor and enhance their quality"; Partially structured: self-evaluation requirements for schools receiving support as part of the Equal Educational Opportunities (GOK) policy.
Belgium (Fr.)	No explicit self-evaluation requirements	1997 Act specifies that each school should draw up a "School Plan" (*Projet d'établissement*) and produce an annual activity report against objectives in the plan.
Canada	School and province/territory	The Atlantic Provinces in Canada are required to conduct School Improvement Planning: Schools are expected to establish a common school vision, long-range 3-to-4 year goals, specific annual objectives, sources of data or indicators, annual reviews and renewal of plans. Schools must produce an annual report for their community and the education authorities evaluating their progress towards goals. Legislation underpins school community (parents) rights to contribute to school improvement planning.
Czech Republic	School	Partially structured (ISCED 2 and 3); highly structured (ISCED 1): Education Act 2005 – schools must conduct self-evaluation and reflect results in an annual school report.
Denmark	No explicit requirements (ISCED 1 and 2); central (ISCED 3)	Partially structured (ISCED 3): although public schools (ISCED 1 and 2) are not required to develop an annual report, this is typically requested by their municipality as part of its legal requirement to produce an annual municipal quality report (*Folkeskole* Act 2006). The municipal report must include nationally specified core indicators.
Estonia	School	Partially structured: Since 2006 schools must conduct self-evaluation using centrally established evaluation criteria. Every 3 years each school must produce a self-evaluation report.
Finland	Local authorities	Education providers evaluate the education they provide, but may decide on the scale, target and implementation of the evaluations.
France	Regional (ISCED 2 and 3); local (ISCED 1)	Partially structured: since 1989 all schools must develop a school plan (*projet d'école ou d'établissement*) with precise goals (in relation to national objectives and curricula) and evaluation procedures to measure the achievement of results. In addition, all secondary education providers must establish a contractual strategic plan for development (*contrat d'objectifs*) with the regional authorities and conduct a self-evaluation providing indicators and evidence.
Greece	None	Not applicable
Hungary	School	Partially structured: schools must have quality management programmes and these should be aligned with local authority quality management programmes (Public Education Act 2006)
Iceland	Central	Partially structured: each compulsory school systematically evaluates the achievements and quality of school activities (Article 35, Compulsory Education Act 2008) with active participation from school personnel, pupils and parents. Each school reports publicly information on its internal evaluation, its connections with school curriculum guide and plans for improvement.
Ireland	Central and school	Since 2012 schools are required to engage in systematic school self-evaluation. They must produce a school self-evaluation report and a school improvement plan annually on aspects of teaching and learning and make a summary of both available to their school community. The 2011 government literacy and numeracy strategy includes a strong focus on school self-evaluation.

Table 6.4 Requirements for school self-evaluation (*continued*)

	Level responsible	Legal framework
Israel	School and central	Partially structured
Italy	None	Not applicable
Japan	School	Unstructured; Highly structured for ISCED 3 pre-voc/voc.
Korea	Provincial/ regional	Highly structured: Primary and Secondary Education Act and the Enforcement Decree (1996). Notably, schools must develop school reports and such documentation is heavily used in external school review processes.
Luxembourg	Central and school (ISCED 1)	Highly structured: The 2009 law for fundamental schools obliges school committees to write and implement a school development plan every four years (describe strong and weak points, define goals to achieve, identify the means to achieve these goals and evaluate the progress they are making). Schools use a pre-defined standard form designed by the Ministry and must define two to five goals related to either "the school as a learning organisation" or "living in the community".
Mexico	None	Not applicable
Netherlands	No explicit requirement	Unstructured: Law on the Supervision of Education (2002) – school's own "quality care" is one of the pillars of the "proportional supervision" approach. Schools must produce an annual school report, a school plan, a school prospectus and an arrangement for complaints (Acts for Primary School and for Secondary School). Schools are obliged to devise a plan for their quality assurance and improvement and must report to parents on results of this plan. Schools must also report to inspectors on student results and progress.
New Zealand	School and central	Unstructured/partially structured: Compulsory audits and self reviews (Education Act 1989). No standard reporting format for annual school plans and reports. National guidelines state that Boards of Trustees together with school principals and teaching staff must develop a strategic plan, maintain an ongoing programme of self-review and report to students, parents and the school community on achievement and progress. Since 2003, schools have been required to produce an annual school plan and report. From 2012/13 schools will need to report student achievement against national standards.
Norway	Central	Partially structured: The Education Act stipulates that schools shall regularly evaluate the extent to which the organisation, facilitation and delivery of teaching are contributing to the objectives laid down in the National Curriculum for Knowledge Promotion.
Poland	Central and school	Partially structured
Portugal	School	Partially structured: 2002 Evaluation Law established internal school evaluation requirements, but no prescribed approach. The Inspectorate ran a project to evaluate school self-evaluation from 2004-2006. One purpose of the second inspection cycle started in 2011 is to validate school self-evaluation practices.
Slovak Republic	Central	Partially structured: Act in 2003 obliges schools to submit a Report on school educational activities, results and conditions. A 2006 decree specifies the content and structure of these reports.
Slovenia	Central	Organisation and Financing of Education Act 2008: schools must conduct self-evaluations annually and report the results to the body that manages each school.
Spain	None	Not applicable
Sweden	School	Unstructured: Schools and municipalities must document quality management, but are not obliged to use a particular format (from the late 1990s until 2010 schools and municipalities had to produce annual quality reports).
Turkey	Central and school	Highly structured
UK (England)	Central	Highly structured
UK (Northern Ireland)	State	Highly structured: Schools must produce a School Development Plan. The Education and Training Inspectorate evaluates and reports on the school's self-evaluation including how effectively it is used to manage and advance self-improvement of the quality of its provision.
UK (Scotland)	State	Highly structured: Standards in Scotland's Schools Act 2000 requires public schools to produce an annual self-evaluation report and a plan for improvement and to report on a range of indicators to their local authority. Approaches to self-evaluation and the effectiveness of self-evaluation and planning for improvement are evaluated as part of the inspection process in all schools.
United States	State, local and school	This varies among states.

Note: Highly structured self-evaluation means that similar activities are completed at each school based on a specific set of data collection tools; unstructured self-evaluation comprises activities that vary at each site depending on the strengths and weaknesses of the school.

Source: OECD Review; OECD (2011).

There are striking differences in how "self-evaluation" is conceptualised in different countries. In the vast majority of countries, this is stated as a requirement for schools to produce a status report on the school's activities or a strategic report on school development, or indeed both. In Sweden it more generally refers to schools having a quality assurance or management system and does not require the production of specific reports (the requirement for specific reporting was dropped in 2010). However, in some systems schools may be required to produce annual reports or development plans, but there is no explicit requirement for school self-evaluation (the French Community of Belgium, Denmark and the Netherlands). This was the case for many years in Ireland, but a specific requirement of self-evaluation was introduced in 2012. From 1998 until 2012 schools were required to produce a school development plan and self-evaluation was conceived as the first step in this process, although not explicitly stated (Irish Department for Education and Skills, 2012).

Although some systems may not have specific legal requirements for schools to conduct self-evaluation, there may be self-evaluation requirements attached to particular policy programmes and funding arrangements. For example, the educational equal opportunities policy in the Flemish Community of Belgium places demands upon schools receiving additional funds to introduce a quality cycle and planning process (Shewbridge et al., 2011a). This quality cycle is similar to school development planning cycles in a number of European countries that are an important feature of school improvement (Creemers et al., 2007). At the end of a three year quality cycle the Flemish Inspectorate of Education examines the school's self-evaluation and makes a recommendation to the authorities on whether or not to continue the funding. In Mexico, over the past 15 years, a number of federal educational programmes have included as a condition for funding a requirement for schools to conduct a self-evaluation exercise and to produce a plan for improvement. Such plans are known variously as "strategic plans for school transformation", "school project" or "strategic plan for school improvement" and currently around 66 000 schools are involved in major programmes (SEP and INEE, 2011).

In the majority of systems specifying self-evaluation requirements, all schools within the system are expected to conduct self-evaluations on an annual basis (although in France and in Scotland within the United Kingdom, the frequency is not specified for independent private schools) (Table 6.5). Evidence from the OECD's Teaching and Learning International Survey (TALIS) in 2008 allows an insight to actual frequency of self-evaluation activities (Table 6.5). Schools are expected to conduct self-evaluations more frequently (more often than once a year) in Israel (secondary schools) and Poland. However, in TALIS, reports from lower secondary school teachers in Poland indicate that such requirements are not implemented in all schools. Indeed, with the exception of the Slovak Republic and Korea, reports from lower secondary teachers on the frequency of self-evaluations undertaken in their schools appear to be in stark contrast to official requirements – in all cases less frequent than would be expected. One can speculate that some of these teachers may be in schools where self-evaluation exercises do not involve them and these remain in the realm of school leadership. Whatever the interpretation, the TALIS data indicate that setting requirements on the frequency of school self-evaluations does not suffice to actually stimulate self-evaluation practices. That said, in three systems where there were no official requirements for self-evaluations in 2008, but there were external school evaluations (Austria, the Flemish Community of Belgium and Ireland), less than 15% of lower secondary teachers reported that school self-evaluation happens at least on an annual basis. In systems specifying self-evaluation requirements, this proportion was at least double (with the exception of Portugal). Interestingly, in both Italy

and Mexico around 50% of lower secondary teachers report that school self-evaluation happens at least once a year, despite the lack of both self-evaluation and external school evaluation requirements.

Table 6.5 Frequency of school self-evaluations

School self-evaluation is required as part of the accountability system	Frequency of school self-evaluations (2009)			Percentage of lower secondary teachers reporting this is at least annual (TALIS, 2007-08)
	Public schools	Government-dependent private schools	Independent private schools	
Canada (Atlantic Provinces)	Annual	Annual	Annual	a
Czech Republic	Annual	Annual	a	a
Denmark	No requirements	Once every three years (as an alternative to inspection)	a	33%
Estonia	Annual	Annual	a	30%
Finland	Unspecified	Unspecified	a	a
France	Annual	Annual	Unspecified	a
Germany	Unspecified	Unspecified	a	a
Hungary	Annual	Annual	a	60%
Iceland	Annual	Annual	Annual	31%
Ireland[1]	Annual	Annual	a	11%
Israel	More often than once a year (ISCED 2 and 3) Annual (ISCED 1)	More often than once a year (ISCED 2 and 3) Unspecified (ISCED 1)	m	a
Japan	Unspecified (ISCED 3 [pre-] vocational)	a	Unspecified	a
	Annual (ISCED 1, 2 and 3 general)	a	Annual	a
Korea	Once every three years	Once every three years (except ISCED 1)	Once every three years (except ISCED 2)	56%
Netherlands	Unspecified	Unspecified	Unspecified	a
Norway	Annual	Annual	Unspecified	41%
Poland	More often than once a year	More often than once a year	More often than once a year	52%
Portugal	Annual	Annual	Annual	19%
Slovak Republic	Annual	Annual	a	85%
Sweden	Annual	Annual	a	a
Turkey	Annual	a	Annual	45%
UK (England)	Unspecified	Unspecified	Unspecified	a
UK (Northern Ireland)	Annual	Annual	Annual	a
UK (Scotland)	Annual	Annual	Unspecified	a
Systems in which school self-evaluation is NOT required as part of the accountability system[2]:				
- External school evaluation is conducted: Austria (12%)[3], Flemish Community of Belgium (14%), French Community of Belgium, Chile and Spain (37%).				
- **No external school evaluation:** Greece, Italy (49%), Luxembourg and Mexico (53%).				

Notes: The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing. (1) Self-evaluation requirements introduced in 2012; (2) Where available, the percentage of lower secondary teachers reporting that self-evaluation is at least annual is given in brackets; (3) Self-evaluation requirements are being introduced. See Table 6.4.

Source: OECD (2011, 2009).

Stimulating self-evaluation for school development, accountability or both

A study among European school inspectorates (SICI, 2003) revealed a lack of official definitions of self-evaluation. Following on from this, an exploratory study of self-evaluation in eight European systems with external school evaluations (Janssens and van Amelsvoort, 2008), identified various informal definitions of self-evaluation and distinguishes two major concepts:

- A process, directly or indirectly aimed at school improvement – ranging from a narrow definition of a "verification or measurement phase within a quality assessment system or school development plan" to a wider definition of a "systematic process, which includes cyclical activities such as goal-setting, planning, evaluation and defining new improvement measures". This process includes the assessment of quality and the judgement and evaluation of learning, teaching and performance.
- A product which is usually perceived as a source for accountability – ranging from a comprehensive document such as self-evaluation forms designed by external bodies to short overviews of self-evaluation results that refer to other source documents.

The importance of this distinction is that school self-evaluation is generally conceptualised as being an internal matter with a firm focus on school development (e.g. Livingston and McCall, 2005; Maes et al., 2002; Yeung, 2011), but specific products of school self-evaluations are often conceptualised for accountability purposes. For example, the product of self-evaluation can be a source of information for the school community. This is an important form of horizontal accountability. Also – to varying degrees in European countries – the product of self-evaluation may feed into external school evaluation processes (see below). Scheerens et al. (1999) identify the strongest interconnection of self-evaluation with external accountability when results of self-evaluation are subject to meta-evaluations by external school evaluation bodies and serve internal and external purposes, and when results from national or district level assessments are fed back to individual schools. Hooge et al. (2012) argue that school self-evaluation providing real insight into school's quality and processes will play a key role in establishing multiple school accountability. Emerging evidence from an ongoing European project indicates that horizontal accountability to stakeholders coupled with clear expectations in external school evaluation are strong determinants of improvement actions at the school level (Ehren et al., 2013).

It can, therefore, be observed that the setting of requirements for schools to produce specific reports – although it may be primarily aiming to stimulate school self-evaluation – introduces an element of accountability to the self-evaluation process. This may hinder the development function of school self-evaluation. However, sharing school evaluation results with school boards/management, parents and students, per se, is associated with school improvement actions.

Articulating external school evaluation and school self-evaluation

School self-evaluation and external school evaluation can be linked in a variety of ways. Alvik (1996) identifies three predominant models depicting the coexistence of self-evaluation and external evaluation in different countries:

- parallel – in which the two systems run side by side each with their own criteria and protocols
- sequential – in which external bodies follow on from a school's own evaluation and use that as the focus of their quality assurance system
- co-operative – in which external agencies co-operate with schools to develop a common approach to evaluation.

The parallel and sequential models imply that external evaluation should dominate the agenda of accountability; the parallel model implies school self-evaluation is solely for the purpose of school development; the sequential model implies both that self-evaluation results are the basis for external evaluation and that the results of external evaluation are expected to feed into school self-evaluation (Kyriakides and Campbell, 2004). As a school system matures there would be a progression from the parallel, through the sequential to the co-operative model, e.g. schools judged to be performing well receive a "light touch" external evaluation (idem). This concept of maturity underlies the push to integrate external school evaluation and school self-evaluation, as schools are assumed to have the will to drive their performance improvement (Barber, 2004). In reality, different aspects of the parallel and sequential models (and less so the co-operative model) are mixed in European systems and no system uses a purely parallel or sequential model (Ehren and Hendriks, 2010). Among the OECD Review countries, New Zealand provides an example of a system close to the co-operative model (see Box 6.3). Overall, New Zealand's evaluation and assessment agenda has been characterised by strong collaborative work and is conceived as a reciprocal learning process (Nusche et al., 2012).

The degree of articulation between external evaluation and school self-evaluation varies across OECD countries. In the majority of systems with both requirements for self-evaluation and external school evaluation, a school's self-evaluation is a component of this external school evaluation process (Table 6.6). With the exception of Iceland and Korea, schools share the results of their self-evaluation directly with external school evaluation bodies. Indeed, many external school evaluation bodies provide access to their quality indicators (e.g. in inspection frameworks) or provide specific quality indicators to guide school self-evaluation (see below). In Portugal, there is an explicit link between external school evaluation and school self-evaluations, but the government does not impose any particular self-evaluation approach, preferring an approach based on diversity and organic growth (Santiago et al., 2012a). The only European systems with external school evaluation and self-evaluation requirements, but where self-evaluation is not a component of external school evaluation are Estonia, Germany and Turkey. However, in Estonia schools share their self-evaluation results directly with the external school evaluation body.

In Iceland, Korea, the Slovak Republic and Scotland within the United Kingdom, a school's self-evaluation results have a high level of influence on the evaluation of school performance (Table 6.6). In Scotland, this also has a high influence over decisions on possible school closure. This is due to the fact that a school's self-evaluation is at the heart of the external school evaluation approach in Scotland (Box 6.3 provides an example of a similar approach in New Zealand). Schools are expected to take responsibility for their quality and demonstrate a clear commitment to continuous improvement (van Bruggen, 2009). Approaches to self-evaluation and the effectiveness of self-evaluation and planning for improvement are evaluated as part of the external evaluation process in all schools. The external school evaluation, therefore, aims to challenge the school's self-evaluation procedures in a way that minimises intrusion to schools, but drives improvement at the national level.

Table 6.6 Links between external school evaluation and school self-evaluation (2009)

	School self-evaluation results are shared directly with[1]:			Influence of school self-evaluation				
	External school evaluation bodies	Higher level education authorities	External audience	Evaluation of school performance	Evaluation of school administration	The size of the school budget	The provision of another financial reward or sanction	The likelihood of a school closure
Component of external school evaluation								
UK (Scotland)	Yes	Yes	Yes	High	High	High	a	High
Czech Republic	Yes	No	Yes	Moderate	Low	Moderate	Moderate	Low
Slovak Republic	Yes	No	Yes	High	High	Low	Low	Moderate
Korea	a	a	No	High	High	a	Low	a
Iceland	a	No	Yes	High	High	None	None	a
Israel	Yes	Yes	Yes	Moderate	None	None	None	None
Poland	Yes	No	Yes	Moderate	Moderate	None	m	None
UK (England)	Yes	Yes	Yes	Low	Low	None	None	Low
France	Yes	Yes	Yes	Low	Low	None	None	None
Portugal	Yes	No	Yes	Low	Low	None	None	None
Canada (Atlantic Provinces)	Yes	Yes	Yes	m	m	m	m	m
New Zealand	Yes	Yes	Yes	m	m	m	m	m
Japan[2]	Yes	Yes	Yes	m	m	m	m	m
Norway	m	m	Yes	m	m	m	m	m
Netherlands	Yes	No	Yes	m	m	m	m	m
Sweden	Yes	Yes	Yes	m	m	m	m	m
Not a component of external school evaluation								
Estonia	Yes	Yes	Yes	Moderate	High	Moderate	Low	None
Turkey	a	a	No	Moderate	Moderate	None	None	None
Germany	No	No	No	Moderate	Moderate	None	Low	None
No external school evaluation								
Hungary	a	Yes	Yes	m	High	None	None	None
Denmark[3]	a	Yes	Yes	m	m	m	m	m
Japan[4]	a	Yes	Yes	m	m	m	m	m
Finland	a	a	No	m	m	m	m	m

Notes: The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing.

(1) The designated group receives school self-evaluation results without having to request them.

(2) Upper secondary pre-vocational and vocational programmes.

(3) Only independent private schools are subject to external evaluation and there are no self-evaluation requirements for these schools.

(4) Primary, lower secondary and upper secondary general programmes.

Source: OECD (2011).

It is of note that not all systems with external school evaluation have explicit requirements for schools to conduct self-evaluations. In the Flemish Community of Belgium school self-evaluation and external school evaluation are deliberately disconnected from each other. In respect of the freedom of education principle, the Inspectorate of Education is not entitled to obtain the school's self-evaluation results, unless the school offers to provide these (Shewbridge et al., 2011a). Although there are no explicit requirements for school self-evaluation in the Netherlands, the Inspectorate of Education evaluates a school's quality assurance policy and this would include a school's

self-evaluation processes if these are in place (Scheerens et al., 2012). In Ireland, a self-evaluation requirement was introduced in 2012 linked to the National Literacy and Numeracy Strategy. This hopes to address concerns that only a low proportion of schools were identified through external school evaluation as conducting robust self-evaluation, despite the fact that the Inspectorate of the Department of Education and Skills has offered schools a model of self-evaluation since 2003 (Irish Department for Education and Skills, 2012).

Box 6.3 Self-review at the heart of school evaluation: New Zealand

New Zealand strives towards a collaborative model of school evaluation where school self-review and external school review are complementary and build on each other. A high level of trust on each side is essential to such a model. Over the past five years the external review body (Education Review Office, ERO) has pushed the agenda of placing school self-review at the core of the school evaluation process. Schools are increasingly seen as responsible for providing their own accountability information, whereas the ERO guides schools toward continuous improvement. The Ministry of Education and the ERO do not prescribe methods for self-review, but provide tools and offer professional development services. The ERO provides guidance documents, where school self-review is conceived of as a rigorous process in which schools systematically evaluate their practice, using indicators as a framework for inquiry and employing a repertoire of analytical and formative tools. It is expected that schools develop understanding of learning progressions, involve students in the assessment and self-regulation of their own learning and analyse assessment data targeted on underachievement.

On the schools' side, there appears to be a commitment to build a data-driven evidence base and to engage in student surveys. The ERO promotes self-review as something embedded in teachers' thinking and practice. While this may be a challenging goal for many schools, at the leading edge there is evidence of schools in which dialogue around achievement data is ongoing and rooted in classroom practice. There is an emphasis on participatory approaches to school self-review, involving both teachers and students in the process. Students have a part to play in evaluating the quality of their school as well as contributing to external review. Including them in this way requires that they are party to the language of assessment and evaluation and that they have the confidence to articulate their views as well as their concerns. The OECD Review revealed exemplary evidence from schools visited that school principals and teachers have taken this issue seriously and have equipped their students with the skills and vocabulary to talk to external visitors on achievement and quality issues. While this may only be practice at the leading edge rather than system wide, the potential for wider engagement is a clear strength.

The ERO has been engaged over the last few years in advocating evidence-informed inquiry, helping schools to engage in that process, and advising on how to use assessment information for improvement and accountability purposes. Dissemination of good practice, reassuring school staff and equipping them with tools of self-evaluation is promoted through workshops. These can serve to demystify self-review and external review and clarify the links between them. Good practice case studies are used as a catalyst for discussion, as illustrations of what effective quality assurance can look like and how it can improve practice, rather than being seen as simply another ministerial demand. The ERO's definition of factors found in effective schools is also disseminated through a series of monographs, highlighting trends, providing commentary and analysis, and pointing to policy implications and system-wide improvements.

Source: Nusche et al. (2012).

Combining the accountability and development functions

In research commissioned by the Dutch Inspectorate of Education, it is argued that (Janssens and van Amelsvoort, 2008): an inherent imbalance in external school evaluation would mean that the external evaluator's judgement lacks independence due to an urge to advise the school, e.g. external school evaluation report includes too many detailed recommendations and suggestions for improvement or the same national body conducts inspections and offers school support functions; and an inherent imbalance in school self-evaluation would be where the accountability demands impede the school development function, e.g. self-evaluations are written for the external school evaluation body and no longer serve the goal of improving education. The researchers find that the position of school self-evaluation in external school evaluation is weaker in national school evaluation systems placing higher emphasis on the development function (Belgium, Denmark, Hesse and Lower Saxony in Germany) and stronger in systems placing more accountability demands on school self-evaluation activities (the Netherlands, and England, Northern Ireland and Scotland in the United Kingdom). The latter group of systems is judged as being "equally supportive of accountability-oriented and improvement-oriented school self-evaluations". The main conclusion is that accountability demands imposed on school self-evaluation generate accountability-oriented self-evaluations, while improvement demands generate improvement-oriented self-evaluations. The authors, therefore, argue that a mixture of a strong position for school self-evaluation in the external school evaluation, transparent external evaluation criteria, and considerable support for schools in steering towards improvement is the most promising combination for bringing about effective school self-evaluation.

Both New Zealand and Scotland within the United Kingdom attach much importance to ensuring that school self-evaluation and external school evaluation use "the same language". Livingstone and McCall (2005) argue that such an approach means "teachers are much more likely to see external inspection in a developmental perspective rather than a judgemental one". External school evaluations may also change the culture in schools towards more formalised and extended processes of evaluating teaching and learning and data analysis (Rudd and Davies, 2000).

Research on self-evaluation in Dutch primary schools indicates the importance of consistency in accountability and improvement policies (Hofman, Dukstra and Hofman, 2009). Among the 939 schools in their study, 81 were linked with information in the Dutch Inspectorate of Education's database concerning official evaluations of the school quality control, quality of the teaching and learning process and the quality of school outcomes. The researchers conclude that "school self-evaluation policies that are strongly driven by both accountability and desire for improvement have a positive impact" (p. 65). Other research conducted in the Netherlands experimented with an approach in which 27 primary schools were free to adopt their own style of self-evaluation and this was then validated during visits from critical friends and the Dutch Inspectorate of Education (Blok, Sleegers and Karsten, 2008). The researchers highlight that training and guidance is crucial for finding a good balance between school self-evaluation and external school evaluation. Such a balance is defined as "a responsive form of accountability" (p. 393).

Procedures

In this section, school evaluation procedures are discussed with respect to the type of reference standards that are used for school evaluation and how evidence is collected against these.

Aspects assessed

International comparison shows a remarkable degree of convergence on the areas addressed during school evaluation (Table 6.7). While limited data are available, these are grouped into three major areas: educational practices; outcomes; and compliance with rules and regulations. Available information on educational practices is restricted to the quality of instruction. With these caveats, the overview in Table 6.7 shows that most countries cover each of these major areas. It is of note that outcomes are restricted to student performance measures and do not include student, parent and staff satisfaction in the United States (although of course there may be varying practices among different States or districts, for example, New York does administer satisfaction surveys to staff, students and parents as part of its school evaluation). Further, in Estonia satisfaction is covered as part of self-evaluation, but not included in external school evaluation. While Norway addresses satisfaction in both types of school evaluation, student performance is only covered in school self-evaluation and not in external school evaluation.

Table 6.7 Areas addressed during school evaluation

(lower secondary schools, 2009)

	Educational practices: quality of instruction		Outcomes: student performance		Outcomes: student, parent and staff satisfaction		Compliance with rules and regulations	
	External evaluation	Self-evaluation	External evaluation	Self-evaluation	External evaluation	Self-evaluation	External evaluation	Self-evaluation
France, Germany, Iceland, Korea, Netherlands, Poland, Slovak Republic and United Kingdom (England)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Portugal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Estonia	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Belgium (Fl.) and Spain	Yes	a	Yes	a	Yes	a	Yes	a
Hungary and Japan	a	Yes	a	Yes	a	Yes	a	Yes
United Kingdom (Scotland) and Canada (Atlantic Provinces)	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Norway	No	Yes	No	Yes	Yes	Yes	Yes	Yes
United States	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Czech Republic	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

Note: The symbol "a" denotes that this is not applicable.

Source: OECD (2011).

In many countries there are attempts to better integrate external school evaluation with school self-evaluation and/or to better target external school evaluation to those schools in most need of improvement. This has led to a new (or more explicit) emphasis on school leadership and on school policies and effectiveness of practices in school self-evaluation. There are different approaches used to this effect, but the underlying aspect is a school's capacity for improvement. There may be an explicit evaluation of the school's capacity to improve or this may be evaluated as part of the judgement on a school's ability to implement policies and practices that lead to improvement, for example, by conducting effective self-evaluations (see Box 6.4).

Box 6.4 Evaluating a school's capacity for improvement

Judging a school's capacity for improvement: Scotland and England in the United Kingdom and the Flemish Community of Belgium

In Scotland, the United Kingdom, a new approach to external school evaluation (Education Scotland, 2011) includes a specific evaluation and report on the evaluated school's capacity to improve (one of three professional judgements: confident; partially confident; not confident). This is a further step in an approach emphasising that the purpose of school evaluation activities is for school improvement. Scotland has also developed and promoted a self-evaluation model for schools including a set of quality indicators for schools to use ("How good is our school?") (HMIE, 2007). One of six key questions in the self-evaluation model is "What is our capacity for improvement?". This is a core aim of self-evaluation activities: "Self-evaluation is forward looking. It is about change and improvement, whether gradual or transformational, and is based on professional reflection, challenge and support." (p. 6). "The emphasis on impact and outcomes reinforces the principle that self-evaluation is not an end in itself. It is worthwhile only if it leads to improvements in the educational experiences and outcomes for children and young people, and to the maintenance of the highest standards where these already exist." (p. 2).

In England, the United Kingdom, the inspectorate provides a clear definition of the school's capacity to improve in a glossary included in each school's inspection reports. It is defined as "the proven ability of the school to continue improving based on its self-evaluation and what the school has accomplished so far and on the quality of its systems to maintain improvement". One of four possible judgements is made on the school's capacity for improvement. Such information is considered in determining the intensity of future inspections.

In the Flemish Community of Belgium, a school's "policy making capacity" is a central concept defined as "the extent to which schools use the available room for policy making to come to a continuous process of retaining or changing their work in order to improve their educational quality and attain both the external and self-imposed objectives" (Van Petegem and Vanhoof, 2009). The Inspectorate does not directly "measure" the policy-making capacity of schools, it is seen as a conditional and relative characteristic: the degree to which a school develops policies to foster student achievement considering its context, resources and student intake. If a school has policy-making capacity, it is expected to improve eventual shortcomings by itself, without help from others. Because policy-making capacities function as a lever for school improvement, they serve as the purpose for the focus of inspection in the preliminary phase and as a discriminating variable between a straightforward negative and a restricted positive recommendation in inspections (Shewbridge et al., 2011a).

Judging a school's self-evaluation capacity: New Zealand

In New Zealand, good self-review capacity is one important aspect in deciding the frequency of external school reviews. Since 2009, the New Zealand Education Review Office (ERO) has adopted a differentiated approach, whereby it reviews: schools with strong performance and self-review capacity every 4-5 years; schools performing well every three years; and schools experiencing difficulty on an ongoing basis over a 1-2 year period. ERO validates the self-review results of schools where self-review practices are well established and investigates these further in the external review where this is not the case (Nusche et al., 2012). A school being reviewed every 4-5 years will convincingly demonstrate that "a school-wide culture of rigorous critical reflection and self review is contributing to sustaining the school's positive performance and continuous improvement" (ERO, 2011). In a school being reviewed every three years (most schools fall into this arrangement) "there is evidence of critical reflection and established processes for conducting and using self review which support improvement". Finally, in a school being reviewed on an ongoing basis "evidence of self-review practices that are helping to lift student achievement and are likely to support school improvement" may influence the duration of the review over the 1-2 year period.

Sources: Education Scotland (2011); HMIE (2007); Van Petegem and Vanhoof (2009); Shewbridge et al. (2011a); Nusche et al. (2012); ERO (2011).

Reference standards used in school evaluation

Legal standards

A school's compliance with rules and regulations is a fundamental aspect of school evaluation. This is typically referred to as administrative or procedural evaluation. International data indicate that across 27 education systems, school compliance with rules and regulations is an aspect evaluated in all systems conducting external school evaluations, with the exception of the Atlantic Provinces in Canada and Scotland in the United Kingdom) (Table 6.7). In Hungary and Japan, where there are no external school evaluations, school compliance with rules and regulations is evaluated as part of required school self-evaluations. In Hungary, the authorities may also occasionally check school compliancy. Although in Scotland in the United Kingdom school evaluation criteria do not include aspects of school compliancy, public schools are expected to submit compliance-oriented reports (OECD, 2011).

In several systems (see Table 6.3), external school evaluations are based on a legal framework in a general education act and would include verification that schools comply with different regulations specified in those acts. For example, in both the Czech Republic and the Slovak Republic, revisions to education acts required that external school evaluation bodies undertake new responsibilities to check the implementation by schools of the content specified in national education programmes in their specific school programmes. The Czech Republic and the Slovak Republic are among 14 systems in which information on the curriculum is collected during external school evaluation (see Table 6.8). In New Zealand, schools are required as part of their self-evaluation activities to submit both a compliance-oriented report (audit) and an evaluative report on their self-review (New Zealand Ministry of Education, 2010). In Sweden, the checking of school compliance with rules and regulations is the major basis of external school evaluation (Nusche et al., 2011b).

Criteria to evaluate the quality of educational processes and outcomes

Several systems have developed a common definition of a "good school" in order to provide a common base for the evaluation of the quality of educational processes and outcomes. Such definitions aim to provide standard criteria to evaluate quality and these are typically underpinned by educational research and evidence of good practice. Effective schools have students that make more progress than expected given social background factors and prior attainment (e.g. Creemers, 2007; Reynolds and Teddlie, 2000) and have processes in place to improve their effectiveness (e.g. Creemers et al., 2007; Harris and Chrispeels, 2006). The characteristics for effective schools are well understood (Sammons et al., 1995) and are broadly common to many national systems and school cultures. They relate to the quality of teaching and learning; the way teachers are developed and helped to become more effective throughout their careers (e.g. Robinson et al., 2008); the quality of instructional leadership in schools (Leithwood et al., 2006) as well as factors concerning the curriculum, vision and expectations, assessment for learning, the rate of progress of students and their educational outcomes. Factors such as these are generally associated with the quality and standards of schools.

Often criteria for school evaluation are presented in an analytical framework comprising: context; input; process and outcomes or results (see Box 6.5).

Box 6.5 Example of areas addressed in school evaluation: Canada and the Flemish Community of Belgium

Prince Edward Island, Canada

Type of indicator	Description	Example in education
Input indicators	exist within the environment of the school and can be considered controllable variables; resources allocated to or consumed by the school	curriculum, years of teaching experience, class size
Context indicators	reflect each student's home experience; information on factors that impact on results that may or may not be in the control of the school	socio-economic status of students, satisfaction levels, ethnicity, demographics
Process indicators	activities associated with the school or what is done at and/or by the school; what the school does to fulfil its mandate	number of classes taught, number of extracurricular activities
Results indicators	benefits for students during and after completing school	student achievement, learning outcomes

Flemish Community of Belgium

<table>
<tr><th colspan="5">Context</th></tr>
<tr><td>Identification</td><td>Situational location</td><td>History</td><td colspan="2">Regulatory framework</td></tr>
<tr><th colspan="5">Input</th></tr>
<tr><td colspan="2">Personnel characteristics</td><td colspan="3">Student characteristics</td></tr>
<tr><th colspan="3">Process</th><th colspan="2">Output</th></tr>
<tr><td rowspan="2">GENERAL
Leadership
Development of school vision
Decision making
Processes and procedures
Quality assurance</td><td>PERSONNEL
Staff management
Professional development</td><td rowspan="2">EDUCATIONAL POLICY
Curriculum
Coaching and counselling
Evaluation</td><td>STUDENT PERFORMANCE
Developmental objectives
Attainment targets</td><td>OUTCOMES
Subsequent education
Job market</td></tr>
<tr><td>LOGISTICS
Infrastructure and equipment
Well-being</td><td>SCHOOL CAREERS
Progress
Enrolment</td><td>SATISFACTION
Staff
Students
Partners</td></tr>
</table>

Sources: Fournier and Mildon (forthcoming); Flemish Ministry of Education and Training (2010).

In Australia, a National School Improvement Tool was developed in 2012 and is based on nine interrelated domains of practice that have been shown to be characteristics of highly effective schools: an explicit improvement agenda; analysis and discussion of data; a culture that promotes learning; targeted use of school resources; an expert teaching team; systematic curriculum delivery; differentiated teaching and learning; effective pedagogical practices; and school-community partnerships.

In Finland, quality criteria were developed in 2009 as a tool to underpin school evaluation in basic education (Finnish Ministry of Education and Culture, forthcoming). Four major areas relate to the quality of structures and include governance, personnel, economic resources and evaluation. Six major areas relate to students: implementation of the curriculum; instruction and teaching arrangements; support to learning; growth and well-being; inclusion and influence; school-home co-operation; and safety of the learning environment.

In the Czech Republic, evaluation criteria used in external school evaluation vary from year to year, although there is a relatively stable coverage of broad evaluation areas (Santiago et al., 2012). For example in 2010/11 and 2011/12 the following areas were included: equal opportunities for education; school education programmes; school management; personnel conditions; material prerequisites; financial prerequisites; effective organisation of education; effective support of personality development of students; partnership; effective support of development of student key competencies; systematic evaluation of individual and group education results of students; and system evaluation of overall results in education.

In Portugal, there is an explicit evaluation of school self-evaluation (Santiago et al., 2012). The school evaluation criteria cover three main evaluation areas: results; provision of the education service; and leadership and management. Leadership and management comprises Leadership, Management and Self-evaluation and improvement. The particular criteria for self-evaluation and improvement include: coherence between self-evaluation and action for improvement; use of results of the external evaluation in the preparation of improvement plans; involvement and participation of the educational community in the self-evaluation; continuity and scope of self-evaluation; and impact of self-evaluation in planning, organisation and professional practices.

Enhancing the transparency and objectivity in external school evaluation

Faubert (2009) finds anecdotal evidence that a lack of clarity of the criteria used in external school evaluation can undermine the legitimacy of the external school evaluation process. School staff may complain about the lack of clarity of the criteria used, and what are perceived as arbitrary statements from the external evaluators. In countries where the standards and criteria used in external school evaluation are not published, there may be a perception that the conditions under which different external school evaluations take place vary significantly. Where external school evaluation sets clear expectations, norms and standards and stakeholders are engaged with and knowledgeable about the external evaluation process, this has significant impact on schools (Ehren et al., 2013).

In New Zealand, the Education Review Office publishes the criteria it uses in external school evaluation. Systematic feedback collected from school principals following an external review indicates higher levels of confidence in the clarity of the review process (New Zealand Ministry of Education, 2010). There is appreciation that the review process and criteria used are open and transparent. In general, there is much less anxiety in schools about the review process compared to when it was first introduced 20 years ago.

In Portugal, feedback collected by the Inspectorate from schools in 2008/09 shows a high appreciation for access to inspection frameworks, external evaluation methodology and its instruments (89%, 82% and 79% of school principals reported this, respectively) (Portuguese Ministry of Education and Science, forthcoming).

Standard school evaluation criteria may also aim to make the different judgements made by external evaluators more objective and transparent, by specifying the different weight and contribution each indicator takes in forming a judgement on school quality (see Box 6.6).

Box 6.6 Indicators used to judge school quality: The Netherlands

In the Netherlands, to assess the quality of primary and secondary schools, the Dutch Inspectorate of Education has established a set of about 45 indicators in five broad categories: Output; teaching and learning process; care and support; quality assurance; law and regulations. Among the 45 indicators is a subset of standard indicators, which play a crucial role in distinguishing between different recommendations made by the Inspectorate.

The standard indicators used in decision rules[1] for primary education are:

- Output
  - Student achievement at the end of primary education is at least at the level to be expected based on the characteristics of the student population in the school.
  - Student results in Dutch language and arithmetic during their schooling are at least at the level to be expected based on the characteristics of the student population in the school.
- Teaching and learning process
  - The learning content for Dutch language and mathematics covers all the school attainment targets as objectives to be achieved.
  - The learning content for Dutch language and mathematics is offered to a sufficient number of students up to the level of Year 8.
  - Schools with a substantial proportion of students classified with language needs provide Dutch language learning content that fits their educational needs.
  - The teachers give clear explanations of the material.
  - The teachers realise a task-oriented work environment.
  - The students are actively involved in educational activities.
- Care and support
  - The school uses a comprehensive system of standardised tools and procedures for monitoring the performance and development of the students.
  - The school carries out the care in a planned way.

The decision rules for judging the quality of primary schools are set as follows:

Weak school – insufficient student achievement results at the end of primary education, plus an insufficient rating on a maximum of one standard indicator in the areas of teaching and learning process or care and support

Very weak school – insufficient student achievement results at the end of primary education, plus an insufficient rating on two or more standard indicators in the areas of teaching and learning process or care and support

Note: (1) For clarity, an abbreviated form of the decision rule is presented.

Source: Inspectie van het Onderwijs (2012).

Developing criteria for school self-evaluation

The development of specific evaluation criteria for school self-evaluation activities has the benefit of engaging the wider school community in the self-evaluation procedure from the start and creating a sense of ownership over the process. The advantage of engaging school principals, teachers, parents and students in creating their own criteria is that it enables discussion and negotiation of criteria which is a valuable process, but it is also a time-consuming process (Kyriakides and Campbell, 2004). Box 6.7 shows an example of a study involving different schools in the development of self-evaluation criteria.

Box 6.7 Developing evaluation criteria for school self-evaluation

A study on how to effectively design and undertake school self-evaluation was commissioned by the National Union of Teachers in the United Kingdom. Over a seven month period, ten schools participated in developing a school self-evaluation framework. There were striking commonalities among the ten schools in terms of the areas they identified to include in the framework, despite the fact schools were in different locations and sectors (primary, special and secondary schools). Participants were asked to identify criteria for "a good school" and as a separate exercise to rate 23 official inspection criteria in use at the time. The first exercise generated six different perspectives (students, teachers, parents, management, support staff and school governors) and showed much overlap in their choice of criteria, although each group's choices reflected different emphases. To a great extent, the criteria chosen by the different groups also reflected many factors that had been identified in research, but also offered some new perspectives and challenged these. Ten areas were identified: school climate; relationships; classroom climate; support for learning; support for teaching; time and resources; organisation and communication; equity; recognition of achievement; and home-school links.

In rating the official inspection criteria again there was strong agreement on criteria related to meeting students' needs (staff understanding and curriculum); conversely, one in three students supported the criterion on promoting moral principles, but these were not selected by any teachers or school governors; the criterion rated most highly by teachers and support staff was "staff working collaboratively towards shared goals"; only a small proportion of students rated this and when asked about it said that they did not recognise it as a feature of their school (p. 63).

The study found that any self-evaluation framework should (p. 73):

1) Have a convincing rationale (why are we doing this?)
2) Reflect the key priorities of the school/authority/national priorities (what is important in this school?)
3) Enable all of the stakeholders to participate (how can we involve everyone who matters?)
4) Allow for the participation of a "critical friend" (how can we ensure a measure of objectivity?)
5) Lead to action/improvement (what do we hope to do with the evidence?).

Source: MacBeath et al. (1996).

Promoting the use of common criteria in external school evaluation and in school self-evaluation

The common coverage of areas addressed in both external school evaluation and school self-evaluation is largely driven by the development of a common set of school evaluation criteria. This is typically by the promotion of the criteria used for external school evaluation or developed by external evaluation bodies for schools to use in self-evaluation. Already in 2001, self-evaluation criteria and models (typically known as "inspection frameworks" or "self-evaluation frameworks") were available to schools in the Flemish Community of Belgium, the Czech Republic, Finland, Ireland, the Netherlands, Portugal, Sweden, Spain and the United Kingdom (Eurydice, 2004). With the exception of Finland and Sweden all these systems had established external school evaluation procedures in place. In France and the Slovak Republic, schools had access to the criteria, indicators and procedures used in external school evaluation. Many European inspection frameworks and self-evaluation frameworks developed by external evaluation bodies are based on school effectiveness research. Further, a European professional collaboration network (the Standing International Conference of Inspectorates, SICI) actively promotes exchanges among different external school evaluation bodies. This research base and professional collaboration means there are common themes in the major areas included in inspection frameworks and a growing consensus across systems on core criteria that can be used to evaluate the effectiveness of a given school.

Importantly, in the interest of promoting a common vision of a "good school", Ministries and external school evaluation bodies typically involve educators and other professionals in developing school quality indicators and criteria. In the case of the Dutch inspection framework this was developed in collaboration with educators and allows a shared understanding of quality in primary and secondary schools (Scheerens et al., 2013). The Korean school evaluation framework was developed with careful attention to educational research on school effectiveness in Korea and internationally, by undertaking original research on effective schools and at the final stages included expert contributions from teachers, school principals and the research community (Kim et al., 2010). In New Zealand, the Education Review Office has developed evaluation indicators that are "underpinned by research, such as the Ministry of Education's Best Evidence Syntheses, and ERO's experience of effective schools" (New Zealand Ministry of Education, 2010; see also Chapter 8).

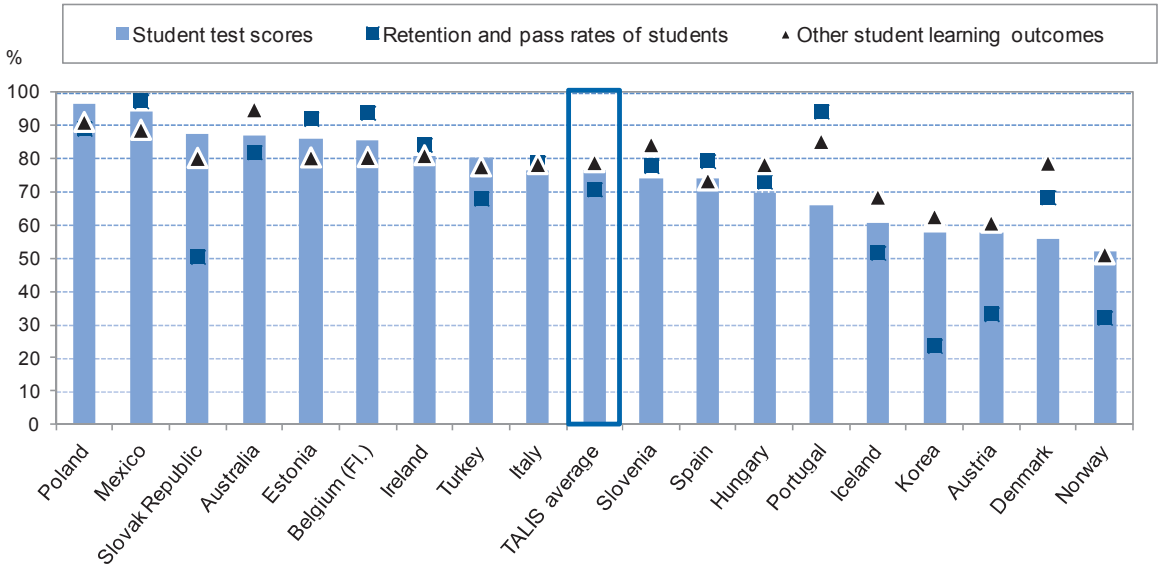
In particular, with regard to frameworks for self-evaluation, it would be remiss not to point out the considerable influence that "the Scottish approach" has had. External school evaluators in Scotland in the United Kingdom are frequently invited to participate in different events across Europe and the external evaluation body (Education Scotland) also receives many international guests (van Bruggen, 2009). Notably, the European Commission funded project on Effective School Self-Evaluation was managed by Scottish external evaluators and an important part of this work focused on the presence of inspection frameworks (SICI, 2003). The Scottish influence has spread beyond Europe also, for example: In Australia, the New South Wales Catholic Education Authority adapted the Scottish inspection framework to develop a rigorous and systemic approach to self-evaluation (Santiago et al., 2011); In Mexico, two major publications that were distributed to schools to promote self-evaluation in 2003 and 2007 were heavily inspired by the quality indicators and other materials developed to support self-evaluation in Scotland (SEP and INEE, 2011).

Student learning objectives

The coverage of outcomes in school evaluation means that student learning objectives should be prominent in reference standards for school evaluation. These may well be incorporated in other reference standards included, for example, in school development plans or as a criterion in external school evaluation. In general, the evidence from TALIS indicates a degree of variation how important student learning outcomes are in both school self-evaluation and external school evaluation (see Figure 6.3). In some systems these seem to be important criteria in school evaluation processes, as reported by the majority of lower secondary teachers. In other systems, significant proportions of lower secondary teachers do not report that student learning outcomes are important criteria in school evaluation.

Schools may be evaluated on the extent to which their students achieve specified learning objectives. In the case of national objectives, these are often measured by proxy with student performance on standardised national assessments or examinations. Standardised national assessments are used in the majority of OECD countries and typically assess student performance in the language of instruction and mathematics (see Chapter 8). Information on student graduation and repetition rates is also used in school evaluation and is often widely available. In the majority of OECD countries participating in TALIS, at least 70% of lower secondary teachers reported that student pass and retention rates are important criteria in school evaluation (see Figure 6.3).

Figure 6.3 Criteria for student learning outcomes used in school evaluation (2007-08)



StatLink http://dx.doi.org/10.1787/888932791400

Note: Percentage of teachers of lower secondary education whose school principal reported that these criteria were considered with high or moderate importance in school self-evaluation or external evaluation. Data are shown for participating OECD countries.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

In the United States, student learning outcomes measured in standardised national assessments form the core of school accountability. Schools must report on student achievement in English language arts, mathematics and on a third indicator that is specified by the particular State (No Child Left Behind Act, 2001). Schools report on the achievement of students overall, plus for different specified groups of students (racial or ethnic groups; special educational needs; economically disadvantaged).

Student performance in national assessments and national examinations also form an important information base for external school evaluations. In France, the mission for external school evaluators (IA-IPR and IEN) is to monitor the implementation of national education policy, laws and regulations in primary and secondary schools in order to contribute to the constant improvement of student knowledge and competencies as defined in programmes for different education levels (Ministère de l'Éducation nationale, 2009). External school evaluators pay particular attention to ensuring and promoting good practices in the follow-up and analysis of results in national assessments and examinations.

Student performance may be included in explicit criteria in external school evaluation frameworks. For example, one of the major quality aspects in the Dutch inspection framework is that "The outcomes of students are at the level that may be expected on the basis of the characteristics of the student population" (Scheerens et al., 2013). The mission for the external school evaluation body in Portugal is to "promote student results and learning progressions, identifying strengths and priority areas for improvement in the work of schools" (Santiago et al., 2012a). This includes specific attention to both academic results (progress of contextualised internal and external student results; quality of success; dropouts) and social results (participation in school activities and acquisition of responsibilities; compliance with rules and discipline; forms of solidarity; impact of schooling on student pathways). In New Zealand, schools will be required to include in their self-evaluation and reporting activities the percentage of their students achieving national standards as of 2012/13 (Nusche et al., 2012).

Instruments and information sources

In collecting evidence for school evaluation, many different methods may be used. This can encompass the use of different tools and instruments designed to collect both quantitative and qualitative information, as well as the direct observation of teaching and learning processes. Further, this may be collected by several different actors.

Administrative reporting by schools

Across OECD countries it is common practice for public schools to submit compliance-oriented reports to education authorities, whether these are at the national, regional or municipal or local level (OECD, 2011).

Administrative reporting by schools includes information on student data in 27 of the 28 systems for which comparable information is available (Table 6.8). Educational authorities typically collect reports from public schools on their facilities and grounds, on teachers' qualifications/credentials, safety issues and closing budget or financial audit from the previous year (OECD, 2011). Among the seven systems where public schools do not report information on the curriculum to the educational authorities, schools in five of these systems report this to their school boards. It is less typical for public schools to report information on governance issues to educational authorities (OECD, 2011). In Austria, the Flemish Community of Belgium and Turkey, school compliancy on governance issues is checked during external school evaluation (Table 6.8).

Table 6.8 Collecting information for compliance-related reporting

	Student data	Teachers qualifications/ credentials	Curriculum	Safety issues	Facilities and grounds	Proposed budget for subsequent year	Closing budget or audit (previous) year	Issues related to governance
Austria	Internet	Internet	Internet	a	Internet	Internet	Internet	Inspection
Belgium (Fl.)	Internet	Internet	Paper/ inspection	Paper/ inspection	Paper/ inspection	No	Internet	Inspection
Belgium Fr.)	a	Internet	a	a	a	a	a	a
Chile	Internet	Internet	Internet	Paper	Paper	Paper	Paper	Paper
Czech Republic	Internet	Internet/ inspection	Internet/paper/ inspection	Paper/ inspection	Internet	Paper	Paper/ inspection	Internet/paper
Denmark	Internet	No	Internet	Internet	Internet	Internet	Internet	No
Estonia	Internet	Internet	Internet	Inspection	Internet	Internet/ inspection	Internet	Internet
Finland	Internet	Internet	Internet	Internet	Internet	Internet	Internet	Internet
France	Internet	a	a	a	Internet	Internet	Internet	Internet
Germany	Internet	Internet	Internet	Internet	Paper	Paper	Paper	m
Greece	Internet	Paper	Paper	Paper	Paper	a	Paper	Internet
Hungary	Internet	Internet	Internet	Internet	Internet	Internet	Internet	Internet
Iceland	Internet	Paper/ inspection	Internet	Paper/ inspection	Paper/ inspection	Paper	Paper	Paper/ inspection
Ireland	Internet/ inspection	Paper/ inspection	Paper/ inspection	Paper/ inspection	Paper	Paper	Paper	Paper/ inspection
Israel	Internet	Internet	Paper/ inspection	Paper	Paper	Paper	Paper	Paper
Italy	Internet	No	Internet	Paper	Paper	Paper	Paper	No
Korea	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection
Luxembourg	Internet	Paper	Paper	Paper	Paper	Paper	Paper	Paper
Mexico	Paper	Paper	a	a	a	a	a	a
Netherlands	Internet	Paper	Paper	Internet	Internet	a	Internet	Internet
Poland	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection
Portugal	Internet/ inspection	Internet/ inspection	Paper/ inspection	Paper/ inspection	Paper/ inspection	Paper/ inspection	Paper/ inspection	Paper/ inspection
Slovak Republic	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet	Internet	Internet/ inspection
Spain	Internet	Internet	Internet	Internet	Internet	Internet	Internet	Internet
Sweden	Internet	a	a	a	a	a	a	a
Turkey	Internet	Inspection	Paper/ inspection	Inspection	Inspection	Paper/ inspection	Paper/ inspection	Inspection
United Kingdom (England)	Internet/ inspection	a	Inspection	a	Inspection	a	a	a
United Kingdom (Scotland)	Internet	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet/ inspection	Internet	Internet	Internet

Note: The terms "paper" or "Internet" denote the method of administrative reporting by schools. "Inspection" denotes that school compliance is checked during external school evaluation. The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing.

Source: OECD (2011), with updated information from the OECD Review.

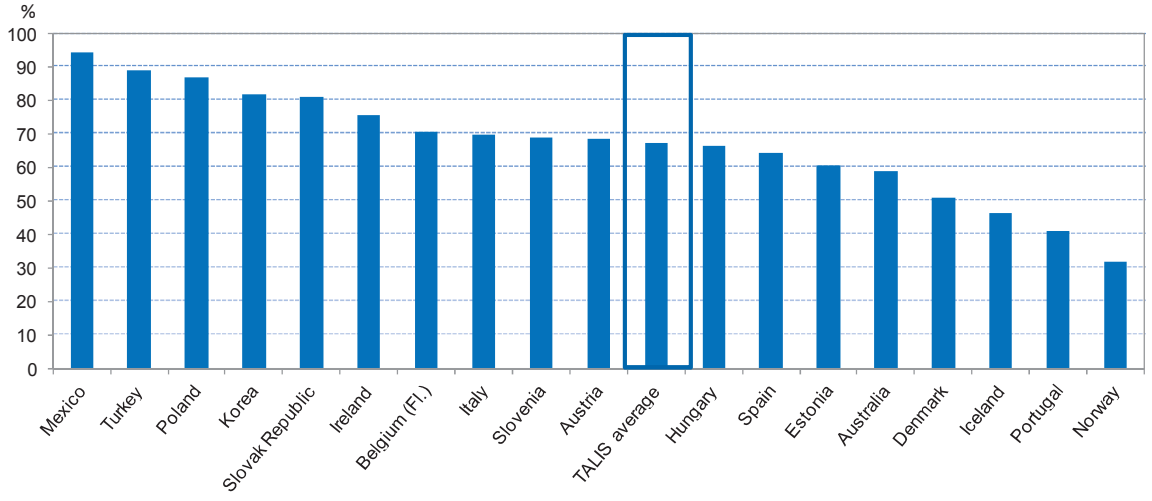
In many countries there are efforts to streamline collection of such information and to ease the reporting burden on schools. Table 6.8 presents an overview of the role of technology in compliance-related school reporting in public schools. Information is collected via Internet-based forms in all countries to some extent, with the exception of Mexico. In some countries, schools have the possibility to submit required information via Internet-based forms in all of the specified areas (Austria, Denmark, Estonia, Finland, France, Hungary, Korea, Poland, the Slovak Republic, Spain and Scotland within the United Kingdom).

In several systems, compliance information is also collected during external school evaluation, notably on the curriculum, teachers' qualifications and the facilities and grounds. In Korea and Poland information is collected during external school evaluation in all of the specified areas (Table 6.8).

Classroom observation

The quality of the teaching and learning process is arguably at the heart of school improvement. Therefore, the direct observation in classrooms of the teaching and learning process should provide key information to school evaluation processes. In systems with school inspections or external reviews, classroom observations are a typical and key part of external school evaluation processes. Here, the emphasis has shifted over the years to an evaluation of teaching quality in the school and not of the individual teacher. However, external school evaluation does have a moderate degree of influence over the evaluation of individual teachers in eight OECD systems and a high degree of influence in the French Community of Belgium, Ireland, Poland, the Slovak Republic, Turkey and England in the United Kingdom (see Table 6.12). Chapter 5 presents an overview of classroom observation for teacher appraisal.

Figure 6.4 Direct appraisal of classroom teaching in school evaluation



StatLink http://dx.doi.org/10.1787/888932791419

Note: Percentage of teachers of lower secondary education whose school principal reported that this was considered with high or moderate importance in school self-evaluation or external evaluation. Data are shown for participating OECD countries.

Source: OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Information collected from lower secondary school principals in the OECD's 2007-08 survey of teaching and learning indicate that classroom observation is accorded relatively less importance compared to other measures of the quality of instruction (OECD, 2009). At least 80% of teachers surveyed were in schools whose principal reported that the direct appraisal of classroom teaching was considered with high or moderate importance in school evaluations in Mexico, Turkey, Poland, Korea and the Slovak Republic (Figure 6.4). In contrast, this was less than 50% of teachers in Norway, Portugal, Iceland and Denmark. In both the Czech Republic and the Slovak Republic the OECD Review found an "open class" culture where the direct observation of classroom teaching was a well established part of school life (Santiago et al., 2012b; Shewbridge et al., forthcoming).

In addition to adequate training, developing and using a set of common indicators for classroom observation can bring more coherence to classroom observations conducted by external school evaluators. For example, an international instrument for teacher observation and feedback (ISTOF) has been developed by educational effectiveness researchers in 19 countries (Teddlie et al., 2006) (see Chapter 5). There has also been an international effort to develop observational instruments for use by external school evaluators. The International Comparative Analysis of Learning and Teaching (ICALT) was a collaboration among European external school evaluation bodies to develop an instrument to observe and analyse the quality of teaching and learning in primary schools. This was developed and piloted by external school evaluation bodies in the Flemish Community of Belgium, Lower Saxony in Germany, the Netherlands and England in the United Kingdom. The study found that the following five aspects could be compared in a reliable and valid way and that these were positively correlated with student involvement, attitude, behaviour and attainment: efficient classroom management; safe and stimulating learning climate; clear instruction; adaptation of teaching and teaching-learning strategies (van de Grift, 2007). The observation instrument was further developed and complemented with a few interview questions (see Table 6.9 for an illustration). The final observation instrument was adopted for use by external school evaluation bodies in the Flemish Community of Belgium, Lower Saxony in Germany, the Netherlands, the Slovak Republic, and Scotland in the United Kingdom (Inspectie van het Onderwijs, 2009).

In the Flemish Community of Belgium, the Flemish Inspectorate of Education has developed a Quality Indicator Model to improve the inter-rater reliability of judgements on the quality of school processes (as specified in the inspection framework) (Shewbridge et al., 2011a). This helps external school evaluators to map out differences among schools in judging the way processes within a school lead to its output – an important part of the external school evaluation approach. The model includes four inter-related categories: Result orientation (drawing up clear and concrete objectives and ways to account for these); Support (staff capacity and material and structural support to achieve objectives); Efficiency (accounting for how school processes contribute to achieving objectives); and Development (attention to continuous development and quality improvement).

Table 6.9 Classroom observation indicators to evaluate the quality of teaching and learning

Rate[1] – Please circle the correct answer: 1 = predominantly weak; 2 = more weaknesses than strengths; 3 = more strengths than weaknesses; 4= predominantly strong.
Observed[2] – Please circle (voluntary) the correct answer: 0 = no, I did not observe this; 1 = yes, I have observed this.

Safe and stimulating learning climate (5 indicators): The teacher…	Rate[1]	Good practice examples: The teacher…	Observed[2]
…ensures a relaxed atmosphere	1 2 3 4	…addresses the children in a positive manner …reacts with humour and stimulates humour …allows children to make mistakes …demonstrates warmth and empathy toward all students	0 1
…shows respect for the students in behaviour and language use	1 2 3 4	...allows students to finish speaking …listens to what students have to say …makes no role-confirming remarks	0 1
…promotes the mutual respect and interest of students	1 2 3 4	…encourages children to listen to each other …intervenes when children are being laughed at …takes (cultural) differences and idiosyncrasies into account …ensures solidarity between students …ensures that events are experienced as group events	0 1
…supports the self-confidence of students	1 2 3 4	…feeds back on questions and answers from students in a positive way …pays students compliments on their results …honours the contributions made by children	0 1
…encourages students to do their utmost	1 2 3 4	…praises students for efforts towards doing their utmost …makes clear that all students are expected to do their utmost …expresses positive expectations to students about what they are able to take on	0 1
Involvement of students (3 indicators): Students…	**Rate[1]**	**Good practice examples: Students…**	**Observed[2]**
There is good individual involvement by the students	1 2 3 4	…are attentive …take part in learning/group discussions …work on the assignments in a concentrated and task-focused way	0 1
…are interested	1 2 3 4	…listen to the instructions actively …ask questions	0 1
…are active learners	1 2 3 4	…ask “deeper” questions …take responsibility for their own learning process …work independently …take initiatives …use their time efficiently	0 1

Note: These are a subset of the observation indicators. Twenty-seven additional teacher-focused observation indicators include: Clear and activating instruction (10); Classroom management (4); Adaptation of teaching to diverse needs of students (4); and Teaching learning strategies (9). There are also four additional student-focused indicators for Reflexivity and discursiveness. The full set of observation indicators is complemented by nine interview questions to assess: Opportunity to learn the minimum objectives (e.g. How many weekly hours are spent on arithmetic?); Monitoring of student progress (How many times a year are the achievements of students tested with standardised tests?); and Special measures for struggling learners (e.g. Does the teacher diagnose the learning problems of students at risk?).

Source: Inspectie van het Onderwijs (2009).

Student performance data

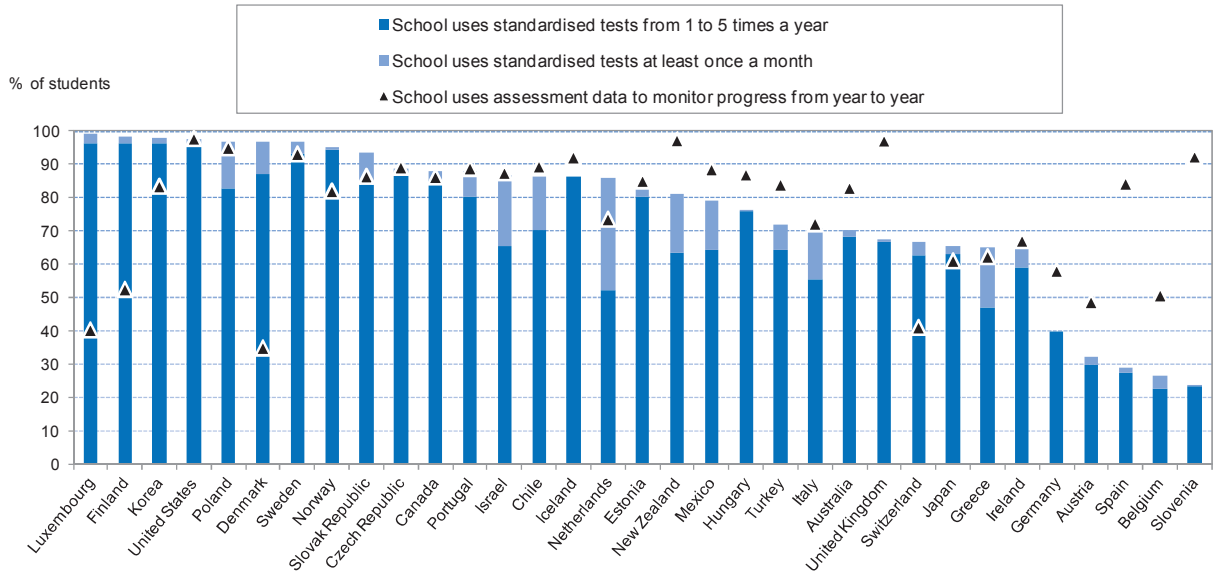
The use of student performance data can make an important contribution to both self-evaluation and external school evaluation. Many systems administer national examinations and national assessments (see Chapters 4 and 8). The use of standardised assessments allows the comparison of different schools on measurements in discrete areas of the curriculum. Assessments that are administered to all students in all schools will provide information that can feed into school reporting systems. In many systems, where

such assessments exist, the results are reported publicly as a form of accountability. In some systems, the results from national assessments are communicated back to schools for their use in self-evaluation activities.

The use of school performance information can increase the efficiency of external school evaluations by helping to target schools that would most benefit from external evaluation. In particular, the use of value added performance information is seen as an ideal complement to the subjective nature of external school evaluations of "what works" as it provides an accurate measure of school performance (OECD, 2008).

Surveys administered to school principals during the PISA 2009 assessment allow a glimpse into the use of standardised tests in secondary schools (Figure 6.5). When answering this question, school principals may have considered not only national assessments, but also commercial tests purchased by schools, or indeed, only commercial tests. The use of standardised tests is clearly a well-established form of collecting information on student performance. In all but five countries over 60% of students were in schools where these were reportedly used at least once a year. It is also clear that secondary schools use other data from other types of student assessment in their self-evaluation activities. In several of the systems where at least 80% of students are in schools reporting the use of assessment data for monitoring, much lower proportions of students are in schools reportedly using standardised tests – notably in Slovenia, Spain and the United Kingdom. Schools in these and several other systems do not always draw on standardised tests for their monitoring purposes. Indeed, there is a weak correlation across the OECD between the use of assessment data to monitor the school's progress and the use of standardised tests in schools (correlation of 0.22).

Figure 6.5 School use of standardised tests and assessment data (PISA 2009)



StatLink http://dx.doi.org/10.1787/888932791438

Note: Percentage of students in schools where the principal reported this happens in the national modal grade for 15-year-olds. Data are shown for OECD countries.

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)* , OECD Publishing, Paris.

Collecting feedback from key stakeholders

Feedback from students, parents, teachers, school leadership and school governing boards can help give a more rounded evaluation of a school's quality. External school evaluations generally incorporate visits to the school. In this case, interviews and discussions with key stakeholders is an important way to collect evidence on how the school is performing. Prior to external school evaluations, external school evaluation bodies may also draw on feedback from parents to help determine the focus and frequency of external school evaluation. For example, in the Netherlands the Inspectorate of Education takes account of parental complaints in its risk assessment when planning external school evaluation activities. The use of questionnaires and surveys may also be an important component of the actual external school evaluation. Surveys may be administered to parents, teachers and staff. For example, in England in the United Kingdom the external school evaluation body offers an online survey for parents and will use this as part of its deliberations in determining the frequency of external school evaluation.

Systems may also develop surveys to collect feedback from key stakeholders and offer these to schools for use in their self-evaluation activities (see Chapter 8). Some examples of different stakeholder surveys are shown in Box 6.8. In some systems, there are also processes to ensure qualitative feedback and evaluation from the school board in annual school reports (see below).

Box 6.8 Collecting feedback from students, teachers and parents

For school self-evaluation

In Norway, an annual survey of all school students in Years 7 and 10 of compulsory schooling is undertaken. This forms an important part of the evaluation of the school system in Norway and the reporting and analysis of national average results have a prominent position in the annual summary report on schooling in Norway. However, results from this are also made available to schools via the School Portal for use in self-evaluation activities. Results are benchmarked to national and regional results, but schools cannot see the results of other schools (Nusche et al., 2011a). Nationally developed surveys for parents and teachers are also available for schools to use in their self-evaluation activities. However, these results are not collected and presented on the School Portal. They are conceived as tools for school self-evaluation to allow triangulation of results from pupils, teachers and parents and to thus provide a richer set of information on that particular school.

In the Flemish Community of Belgium, there are no specific standardised surveys used for the collection of feedback from stakeholders. While this general approach to resist standardised surveys is supported by students, the secondary student organisation has developed a suite of possible self-evaluation tools that teachers can use to get feedback from their students on their perception of the teaching and learning experience. This aims to strengthen student voice while remaining non-threatening to teachers, as the major purpose is to provide constructive feedback for improvement (Shewbridge et al., 2011a).

For the reporting of comparative measures on school performance

In the United States, the city of New York (NYC) systematically collects information from parents, teachers and students. The "NYC School Survey" is administered each year to all students in Years 6 to 12. This collects information on the school's learning environment, including questions on school safety and respect, academic expectations, student engagement and communication among the school community. Results feed into external accountability measures (the school's progress report) and form between 10-15% of the school's overall assessment. Average results for NYC are compiled and reported to the public and can be used for school benchmarking, thus also feeding into school self-evaluation activities.

Collecting feedback from stakeholders on the external school evaluation process is also an important way to improve external school evaluation capacity (see below).

Some systems allow a mechanism for the school to verify and comment on the content of the external evaluation report before it is published. It is typical practice for external school evaluators to discuss briefly an overview of their findings at the end of the actual visit (Faubert et al., 2009). However there may also be possibilities for schools to provide written comment on the report prior to publication.

In the Czech Republic, the content of the external school evaluation report is discussed between school inspectors and the school principal (Santiago et al., 2012). The school principal confirms through his/her signature that the report/protocol has been discussed. The school principal may submit his/her comments on the external school evaluation report to the Czech School Inspectorate (within 14 days after it was submitted) or objections to the protocol (within 5 days after it was submitted). These comments are included in the final report which will be sent to the organising body and the school board. The external school evaluation report is published on the Czech School Inspectorate website and is kept for a period of ten years.

Compiling an evidence base for external school evaluation

There are various approaches to compiling an evidence base for external school evaluation. In addition to conducting observation of teaching and learning or administering special survey questionnaires to stakeholders, external evaluators may draw on centrally compiled statistical information on schools, or review extensive documentation on the school and developed by the school, or conduct interviews with members of the school.

In the Flemish Community of Belgium, the Inspectorate of Education uses information in a Data Warehouse system provided by the Ministry of Education and Training (Shewbridge et al., 2011a). Data are compiled from compliance reports submitted by schools. The Inspectorate of Education constructs an individual school profile including indicators on output, input and context over a six year period. Each school is benchmarked against a group of comparable schools. This supports the Inspectorate of Education in its preliminary risk analysis to decide the focus of external school evaluation. In addition, during the actual external school evaluation, to assess the context, input and process factors, the investigation draws most heavily on document analyses, interviews with teachers, students and school leadership, as well as classroom observations. A challenge for external school evaluation is to have adequate information on student outcomes at the risk analysis phase. Further, schools are not obliged to share their self-evaluation results with external school evaluators during the external school evaluation.

In the Czech Republic, the approach to external evaluation is designed to be evidence driven (Santiago et al., 2012b). The provision of a data profile for an inspection team, provided by the Institute of Information on Education, offers outcome information, aids efficiency by allowing the team to focus its attention on key issues and can help to benchmark and contextualise judgements. Similarly, documentation is sought and analysed as a key part of evidence gathering and a sample of stakeholders is interviewed in the course of the external school evaluation. As a result, evaluation teams have a wide body of evidence upon which to base their judgements. Moreover it appears that the external school evaluation process seeks to take into consideration contextual factors that

influence performance such as school type and location, kinds of students served, although this is not done systematically.

In Portugal, each externally evaluated school completes a "presentation" document. This provides information mapped to the areas analysed in external school evaluation and linked to its self-evaluation (Santiago et al., 2012a). The General Inspectorate of Education and Science provides the external school evaluators with a statistical profile of the school, including performance data from national assessments and national examinations, statistics on grade repetition, and data on the demographic and social characteristics of the student population. External school evaluators also review documents such as the educational project, the curricular project, the plan of activities, the internal regulations, and the self-evaluation report. Another major instrument is panel interviews with the representatives of the educational community: school leadership, teaching and other staff, students, parents/guardians, and the local authority, selected according to pre-specified criteria. Triangulation across different sources of evidence is used to promote reliability. A new instrument is the administering of questionnaires to students, parents/guardians, teachers and non-teaching staff on their satisfaction and analysis of the school results. External evaluators also observe the school facilities, including the areas for instruction, but do not directly observe teaching and learning.

Capacity

Capacity for external school evaluation

A crucial concern in external school evaluation is to ensure the legitimacy of the external school evaluators (typically known as "reviewers" or "inspectors"). The typical recruitment of external evaluators is from the education sector. External evaluators are most likely to have experience in education or teaching, to be recognised as having in-depth expertise, to be former successful practitioners and to be able to guide and support others in the process of school development. However, there is a tension here of attracting well-experienced educators out of the education sector and thereby lessening the quality within the school system.

It is not always straightforward to recruit external evaluators. For example, in the Slovak Republic it is proving difficult to attract people into the State Schools Inspectorate in particular regions, often due to the lack of additional financial incentive to leave the education sector (Shewbridge et al., forthcoming).

In Korea, a practical challenge has been to secure credible and independent external school evaluators (Kim et al., 2009). Evaluators are typically school principals, school deputy principals, educational supervisors and researchers from within a particular district, so they are credible as evaluators, but have ties with the district and so impartiality may be a concern. Researchers point to the need for higher quality training programmes to address these concerns.

In Sweden, all external school evaluators are full-time civil servants (Swedish Ministry of Education and Research, 2010). The majority of evaluators have a professional background in education and some have experience from working as senior administrators in a municipality or have been researchers in pedagogy and involved in teacher professional development. The external evaluation body (the Swedish Schools Inspectorate) has also recruited individuals trained in law or social sciences and researchers and analysts from various disciplines. This is a strategy to broaden the knowledge base and experience within the Swedish Schools Inspectorate. All external evaluators must have a university education

or equivalent and broad knowledge and experience in their professional field. All new recruits begin a six month probationary period upon employment.

In some systems, reviewers from outside the education sector are used in combination with full-time external evaluators (see Box 6.9). Combinations of this kind can provide reassurance to those being evaluated about the competence and objectivity of teams by bringing different expertise and perspectives to bear during the evaluation process. Full-time external school evaluators develop techniques of evaluation which are specific to this type of work while the "outsider" members of a team can be selected for their own particular expertise and credibility.

In Portugal, external school evaluation comprises a team of three members, comprising two full-time members of the General Inspectorate of Education Science (IGEC) and an external member chosen by higher education or education research institutes (Santiago et al., 2012a). The external team member is usually drawn from a higher education institution.

Box 6.9 Recruiting senior educators to join external school evaluation teams, Northern Ireland, United Kingdom

In Northern Ireland, the Education and Training Inspectorate (ETI) recruits "associate assessors" from among senior staff in schools (e.g. school principals, deputy principals or senior teachers) to participate the external evaluation of individual schools. ETI recruits associate assessors via public advertisement and an interview process. Selected individuals join a pool of associate assessors and can be invited to join an external school evaluation team on an individual school inspection. Normally an individual will not be involved in more than two external school evaluations each year. Associate assessors receive training from the ETI and are introduced to the procedures and performance indicators used in external school evaluation.

This strategy has two objectives: first, it is hoped that the experience of involvement in assessing quality in another educational establishment will help to develop the individual's capacity to monitor, evaluate and improve the provision in his/her own school; second, the presence in the team of someone coming directly from the school context adds a dimension which can help to develop the ETI's awareness of the current perspective of schools.

Source: Department of Education, Northern Ireland (forthcoming).

Upon recruitment, external school evaluators typically follow a specialised training programme on the techniques of external school evaluation (Faubert, 2009).

Aligning external evaluation capacity with the chosen approach to external school evaluation

Governance decisions to define the overall approach to external school evaluation impact the capacity requirements in external evaluation bodies. The OECD Review noted that New Zealand's decision to shift to a system of differentiated external reviews has been accompanied by an investment by the Education Review Office in training its staff in how to handle and interpret evidence from school self-review. This has also been used to stimulate capacity building at the school level (see Box 6.3). Similarly, the Flemish Inspectorate of Education has recently introduced a "differentiated approach" to external school evaluation and is on a "learning path of continuous reflection, refinement and improvement" (Flemish Ministry of Education and Training, 2010). The Flemish

Inspectorate of Education offers generic training to all external school evaluators on the "differentiated approach" and seeks regular feedback from stakeholders via questionnaires and conferences to feed into consideration on how to adjust and refine the new approach to external school evaluation. Hong Kong-China has also invested heavily in retraining external reviewers to shift from an inspection mindset to a review approach, in which external reviewers are conceived as mediators who encourage and support schools to speak for themselves (MacBeath, 2008, p. 395).

In Sweden, external school evaluation is carried out in two major forms: regular supervision and thematic quality reviews (Swedish Ministry of Education and Research, 2010). During regular supervision the main focus is on legality; the purpose is to ensure the right of each individual in relation to the Education Act and regulations that apply. Similarly, ad hoc complaints received by parents are always investigated on a legal basis. In this context, the Swedish Schools Inspectorate has recruited individuals trained in law.

In Korea, there has been a clear shift in policy to introduce a system of whole-school external evaluation, in which school quality is evaluated against a national quality indicator framework (Kim et al., 2010). While this has the advantage of giving a more rounded evaluation than simply relying on school performance measures, it requires much expertise from external evaluators. Training programmes are offered at the national level by the Korean Educational Development Institute and comprise a mixture of lectures and workshops on the legal basis, role and rights of related organisations, basic evaluation plan, interpretation and practical application of indicators, evaluator ethics, and guide to writing evaluation reports. There has also been increased support at the local level to help evaluators practice indicator application. However, research has pointed to the need to increase the offer of training (Jung et al., 2008).

In France, a broad cross-section of stakeholders shares the opinion that external whole-school evaluation would be difficult to introduce due to inadequate external school evaluation capacity and a lack of specialisation in the necessary skills and competencies to undertake this (Dos Santos and Rakocevic, 2012). During the 1990s, certain school organising bodies (*académies* in Paris, Lille, Rouen and Toulouse) implemented whole-school evaluations. These were conducted over a series of years, but proved to be time consuming and demanding in terms of human resources and were not always appreciated by school principals, so they were abandoned. A similar attempt was made in the 1990s by the national inspection (IGAENR) to introduce whole-school evaluations resulting in reports on main school features, strengths and weaknesses sent to the school and its organising body (*académie*), but such reports were not followed up and were therefore abandoned.

Refining and improving external school evaluations

As noted above, adopting a principle of transparency in the methodology and evaluation standards used in external school evaluation can increase the legitimacy of the external evaluation process in the minds of those being evaluated. Also, the transparency of publishing external evaluation reports for individual schools has been found to bring more coherence to the form and content of these (see below).

Another way to heighten the legitimacy of external school evaluators is to ensure that the evaluators are themselves evaluated (Faubert, 2009). This can provide valuable information for improving the capacity of the external evaluation body to conduct objective and impactful evaluations. Most countries with external school evaluation ensure regular discussion of approaches and instruments used in external school evaluation, often under

the supervision of a senior member of the external evaluation body (Faubert, 2009). In order to judge the effectiveness of external school evaluation procedures, many external evaluation providers collect feedback from schools and other stakeholders on their experience with the external evaluation process. The external evaluation procedures may also be evaluated during national audits. Such processes can help validate the external evaluation procedures in place and improve their effectiveness and impact and may be particularly useful when introducing new external evaluation procedures.

In New Zealand, feedback from school principals on their experience with the external review process is systematically collected (New Zealand Ministry of Education, 2010). There are also other channels for stakeholders to feedback concerns or possible directions for future external school reviews: the Education Review Office has a Public Affairs section to answer individual concerns and holds a variety of public meetings and speaking engagements; there is an official complaints procedure; and individuals or groups can lobby their Member of Parliament.

In Ireland, the Department of Education and Science may commission an independent customer survey to seek views on the external school evaluation process. The Department of Education and Skills may also be subject to national auditing processes, for example a recent review praised the work of the external school evaluation body and in particular its evaluation work in schools (Department of Public Expenditure and Reform, 2012). Similarly, in Sweden an audit of the external school evaluation process was conducted by administering a questionnaire to local politicians, civil servants, school principals and teachers in 38 local authorities and by conducting a few case studies and found positive feedback on school external evaluation (Ekonomistyrningsverket, 2006).

In the Netherlands, the Dutch Inspectorate of Education conducted evaluation studies in 2010 to judge its work in general, but also to seek feedback on the new approach of "risk based inspection". On a number of criteria schools reported positive experience with the risk based approach (Inspectie van het Onderwijs, 2010a, 2010b).

Capacity for school self-evaluation

The OECD Review has revealed a common concern among countries on the variation among schools in their capacity to undertake self-evaluation.

A recent evaluation of secondary education in the Flemish Community of Belgium identified the policy-making capacities of schools as an important improvement challenge (Commissie Monard, 2009).

In Austria, teachers report that additional time is required to work on self-evaluation and the lack of additional resources is a major barrier to its implementation (Specht and Sobanski, 2012). School principal reports in PISA indicate that students and parents show positive attitudes to self-evaluation, but that teacher unions are perceived as a hindrance to school self-evaluation activities (Haider, 2006).

In the Czech Republic, research on school self-evaluation activities has revealed that these lack a systematic, coherent and purpose-specific approach (Vašťatková and Prášilová, 2010), but that as schools gain experience in self-evaluation, they see this as a more meaningful activity (also Chvál et al., 2010). There are significant differences among schools in self-evaluation capacity, with a large proportion of schools needing support and the majority of schools demanding validated self-evaluation instruments (Chvál et al., 2010).

In Norway, the tradition of school self-evaluation directly related to the school's own development has been developing since the 1970s. By 2000, almost half of all Norwegian schools and municipalities had developed systematic forms of school evaluation, but it has proved challenging to engage the remaining schools and local authorities in this type of quality assessment (Roald 2010). Oterkiil and Ertesvag (2012) highlight research that shows the failure for evidence-based and nationally supported programmes to be successfully implemented in some Norwegian schools and argue that the key is schools' readiness and capacity to improve.

In the Netherlands, researchers conducted a survey on school self-evaluation activities in 939 Dutch primary schools and classified schools into clusters according to the intensity of self-evaluation activities and accent on accountability and improvement: 8% of schools take hardly any self-evaluation activities for accountability or improvement; 33% of schools have average self-evaluation activities paying some attention to accountability and improvement; 30% of schools undertake advanced school self-evaluation with highly implemented accountability and improvement measures; 29% of schools undertakes above average accountability measures, but low levels of school improvement (Hofman, Dukstra and Hofman, 2009).

A European project in 2001-03 identified a cocktail of elements that contributed to effective self-evaluation, including national and local support for self-evaluation, strong leadership, engagement of key stakeholders in the process and strong staff commitment to school self-evaluation (see Box 6.10).

Box 6.10 Effective self-evaluation

The Standing International Conference of Central and General Inspectorates of Education (SICI) ran a project on "Effective School Self-evaluation" from 2001 to 2003, including 39 schools across 14 different European inspectorates. Clearly, part of the aim of this project was to identify how to best marry the external element of inspection with school self-evaluation practices. However, the final report (SICI, 2003) also presents case studies showing examples of schools in the project with good school self-evaluation practices. The project identified the following common elements among schools with "very effective self-evaluation":

- strong leadership
- school aims which were shared and clearly understood by all key stakeholders
- engagement of key stakeholders in self-evaluation and improvement activities
- well set out and clearly communicated policies and guidelines
- self-evaluation activities that focused on learning, teaching and improving outcomes
- strong staff commitment to self-evaluation
- monitoring and evaluation processes that were systematic, rigorous and robust
- well planned action to develop and improve provision
- a beneficial balance between external support and challenge from local authorities and/or national Inspectorates and internal quality assurance
- a generally strong infrastructure of national or local support for self-evaluation as a process.

Source: SICI (2003).

Offering national or local support for school self-evaluation

Strong national or local support for school self-evaluation has been identified as an important element in ensuring effective school self-evaluation processes (see Box 6.10). Table 6.10 presents an overview of the major national supports for self-evaluation offered in OECD countries. Specific training can be offered to school principals and teachers in areas such as the use of evaluation results, classroom and peer observation, the analysis of data and the development of improvement plans. Training and conferences on self-evaluation activities are offered in a number of countries. Examples of other types of national supports for self-evaluation include guidelines for self-evaluation and the development of school improvement plans, tools for evaluation and data analysis, including the feedback of performance information from national assessments. Box 6.11 shows an example of a comprehensive centrally developed support package for schools in Scotland in the United Kingdom.

Table 6.10 National initiatives to support school self-evaluation

	National support for school self-evaluation
Australia	School level results from national assessments are available to schools via *MySchool* public website. Western Australia, Victoria and New South Wales have developed online information systems to support schools and in particular to feed back results from national assessments in a way that allows schools to analyse these more efficiently. A national school improvement tool was developed in 2012 for schools to self-evaluate on nine major characteristics research has identified in effective schools.
Austria	Quality in Schools (QIS) project Internet platform supplies schools with information and tools for both evaluation and data analysis, and provides a forum for the presentation of results
Belgium (Fl.)	New possibility for schools to conduct student tests as administered in the national sample assessment and to receive benchmarked feedback; feedback reports given to all schools participating in national and international assessments.
Czech Republic	Schools are free to choose self-evaluation criteria, but the Ministry and the Inspectorate provide guidelines, selected models and examples of good practice ("On the road to quality" project).
Denmark	The Quality and Supervision Agency runs an Evaluation Portal with online tools and resources for school evaluation and in collaboration with the Danish Evaluation Institute offers voluntary training sessions for school principals and teachers.
Finland	In 2010 the Ministry of Education and culture devised national quality criteria for basic education with a view to facilitating school self-evaluation and quality enhancement. The Education Evaluation Council produced evaluation guides and methodological publications and disseminates good evaluation practices.
France	Secondary schools have been equipped with dashboards of performance indicators. Data come mostly from the centrally developed application Support for School Piloting and Self-Evaluation (APAE, *aide au pilotage et à l'auto-évaluation des établissements*). The indicators are mostly related to school population, financial and student achievement data. In the special education sector (*éducation prioritaire*), a Support Tool for School Piloting (OAPE, *outil d'aide au pilotage de l'établissement*) is currently being developed to help school teams self-evaluate their activities in order to collectively increase the "school effect".
Hungary	National Centre for Assessment and Examination in Public Education issues guidelines on organisation and methods for self-evaluation and organises training in quality development.
Iceland	Ministry publishes guidelines and offers training to teachers. New curriculum guidelines will include special evaluation guidelines for schools.
Ireland	Strengthened support in 2012 includes: Guidelines for School Self-Evaluation in primary and secondary schools; a dedicated school self-evaluation website; Inspectorate support for all schools and teachers; and seminars for school principals which are organised by the professional development service for teachers. In 2003 the Inspectorate developed two frameworks for self-evaluation in primary and secondary schools (Looking at our schools). Since 1998, professional development for teachers offered in context of School Development Planning.
Korea	MPOEs provide guidelines for schools to complete self-evaluation reports by providing evidence of their educational activities and outcomes for common evaluation indicators in the national evaluation framework. MPOEs provide training to senior school staff on how to prepare for school self-evaluation activities.
Luxembourg	Ministry accompanies schools in their school development planning by offering data, assessment tools, advice, training and analytical expertise and analysing data. Methodological support is offered to schools throughout the process of drawing up and implementing their School Development Plan by the central Agency for the Development of Quality in Schools (ADQS).

Table 6.10 National initiatives to support school self-evaluation (*continued*)

	National support for school self-evaluation
Mexico	Self-evaluation guidance developed since the early 2000s, including an adaptation of the Scottish evaluation and quality indicator framework (2003) and a publication on key features of the top performing schools (2007). Further a collection of guides, support materials and instruments for self-evaluation was distributed to all primary and secondary schools in 2007 (System for School Self-evaluation for Quality Management). The National Testing Institute (INEE) also develops a series of applications for use in self-evaluation, e.g. tools for evaluating the overall school, the school environment, school staff, etc.
Netherlands	Ministry subsidised two national projects to develop quality management systems in secondary schools (1999-2005) and in primary schools (2001-06). Secondary School Council (2010) in co-operation with the Ministry has developed an online information system with quantitative and qualitative information on individual schools with benchmarking data that can be used for self-evaluation and horizontal accountability.
New Zealand	Education Review Office provides support tools and training for school self-review and improvement, suggesting a cyclical approach and providing a framework for success indicators (same as those used in external reviews).
Norway	A national "School Portal" presents benchmarked outcome data for school owners and schools, plus basic demographic and resource data. National template for school reporting includes mandatory and suggested indicators. These are part of the National Quality Assessment System (NKVS) introduced in 2004.
Portugal	General Inspectorate of Education and Science provides school self-evaluation support materials on its website and organises seminars on good practices in self-evaluation. The Inspectorate's "School Presentation" instrument guides schools in how to present their own evaluation at the start of the inspection process.
Slovak Republic	The Ministry published methodological guidelines in 2006 to help schools produce their self-evaluation reports. There are no defined criteria. There are three ongoing national projects on school self-evaluation (cofounded by European Social Funds).
Slovenia	Various projects to support self-evaluation via National School for Leadership and National Education Institute. Ongoing project (co-funded by European Social Funds, 2008-14) to develop a system of quality assessment and assurance focused primarily on self-evaluation, combined with external school evaluation and quality indicators.
Sweden	The National Agency for Education provides a school self-assessment tool (Assessment, Reflection, Evaluation and Quality, BRUK) with indicators on the national curriculum and syllabi, plus a tool to plan and assess internal quality improvement. Many privately developed tools also available to schools.
United Kingdom (England)	A national self-evaluation form is provided for schools – this is used by schools prior to school inspections. An analysis tool Reporting and Analysis for Improvement through School Self-Evaluation (RAISEonline) is provided for use by schools, local authorities, inspectors and school improvement partners.
United Kingdom (Northern Ireland)	The Education and Training Inspectorate has developed in collaboration with schools and practitioners a set of quality indicators (Together Towards Improvement) which is promoted for use in school self-evaluation. Other tools and guidelines have been developed to support both whole-school evaluation and evaluation in specific subjects, e.g. "Evaluating English".
United Kingdom (Scotland)	Framework for school self-evaluation (How good is our school?) includes quality indicators in five key areas. Education Scotland website also provides a range of self-evaluation materials and good practice examples. Education Scotland runs good practice conferences on different themes.

Source: OECD Review.

It is worthy of note that efforts to build school capacity for self-evaluation have been undertaken by many countries already for a number of years. In 2001, nearly all European Union members had some form of support for self-evaluation in place, the most common support being training for self-evaluation (e.g. in Austria, Belgium, Finland, Germany, Hungary, Iceland, Ireland, Norway, Poland, Slovenia, Spain and the United Kingdom) (Eurydice, 2004). Self-evaluation frameworks and models were available in the Flemish Community of Belgium, the Czech Republic, Finland, Ireland, the Netherlands, Portugal, Sweden, Spain and the United Kingdom. With the exception of Finland and Sweden all these systems had established external school evaluation procedures in place. In France and the Slovak Republic schools had access to the criteria, indicators and procedures used in external school evaluation. However, there is continued concern to strengthen school self-evaluation capacity. For example, in New Zealand, heightened priority has been given to building school capacity for self review (New Zealand Ministry of Education,

2010). The Education Review Office assessed in 2007 that around half of schools were considered to be conducting their self review well, that is, they were using assessment information well to inform teaching and school decision making (ERO, 2007). Within the United Kingdom, external school evaluation in Scotland has sought to build capacity by providing more extended engagement and support to schools most in need of this and by creating more scope for inspectors to work directly with school staff during external school evaluation (van Bruggen, 2009).

Box 6.11 Centrally developed tools for self-evaluation in Scotland, the United Kingdom

The external evaluation body in Scotland (Education Scotland) has developed a central web-based resource which provides schools and school managers with a comprehensive set of tools which they can use to structure effective school-level evaluation. This resource, known as Journey to Excellence has grown and developed over two decades and can be traced back to the publication of *How Good is our School?* in the late 1980s.

The complete Journey to Excellence package now includes the following parts:

- Part 1: *Aiming for Excellence*; explores the concept of excellence, what is meant by "learning" and "barriers to learning" and introduces ten dimensions of excellence.
- Part 2: *Exploring Excellence*; explores the ten dimensions in detail, giving practical examples from real schools which show the journey from "good" to "great".
- Part 3: *How Good is our School?* and *The Child at the Centre* present sets of quality indicators for use in the self-evaluation of schools and pre-school centres respectively, along with guidance on their use.
- Part 4: *Planning for Excellence* provides a guide for improvement planning in schools and pre-school centres.
- Part 5: *Exploring Excellence in Scottish Schools* consists of an online digital resource for professional development containing multi-media clips exemplifying aspects of excellence across a wide range of educational sectors and partner agencies. It also contains short videos from international education experts and researchers.

Plans are underway to enhance the resource further with new resources to support schools in the process of developing long-term strategic thinking and managing major change in a school context.

The package is very widely used by schools and by all Scotland's 32 local authorities and most independent schools. The school quality indicators at the heart of the package are also used by external school evaluators for external review of schools. They were built on the criteria developed for external school evaluation and they are regularly refreshed and updated on the basis of developing understanding of the characteristics of effective practice.

Source: HMIE website (*www.hmie.gov.uk/generic/journeytoexcellence*).

Ensuring leadership of school self-evaluation activities

The key role that strong leadership plays in ensuring effective school self-evaluation has also been highlighted (see Box 6.10). Research internationally has shown that school leadership focused on goal-setting, assessment, appraisal and evaluation can positively influence teacher performance and learning environments (Pont et al., 2008). Evidence

from TALIS indicates that if school principals adopt a more pronounced instructional leadership style, teachers are more likely to engage in collaborative activities with their colleagues (this is the case in more than a quarter of the TALIS countries) (OECD, 2009). In Ontario, Canada, professionals have developed a set of competencies for school principals related to school self-evaluation (Box 6.12).

Box 6.12 School self-evaluation related competencies for school leadership: Ontario, Canada

The Institute for Educational Leadership (IEL) in Ontario is a virtual organisation made up of a partnership of representatives from Ontario's principals' and district officers' associations, councils of school district directors, and the Ministry of Education. Its purpose is "to further develop education leadership so as to improve the level of student achievement in Ontario's four publicly funded education systems. One of IEL's five practices and competencies within its research-based leadership framework for school principals and deputy principals is "leading the instructional program", described as: "The principal sets high expectations for learning outcomes and monitors and evaluates the effectiveness of instruction. The principal manages the school organisation effectively so that everyone can focus on teaching and learning." Among a number of practices outlined to achieve this are: ensuring a consistent and continuous school-wide focus on student achievement; using data to monitor progress; and developing professional learning communities in collaborative cultures. Associated skills include that the school principal is able to access, analyse, and interpret data, and initiate and support an enquiry-based approach to improvement in teaching and learning. Related knowledge includes knowledge of tools for data collection and analysis, school self-evaluation, strategies for developing effective teachers and project management for planning and implementing change.

Source: Ontario Institute for Educational Leadership website, *www.education-leadership-ontario.ca/content/framework*.

In the Czech Republic, a study of external support offered to school principals on the self-evaluation process in 12 primary schools in Prague showed that this had a positive impact on school principals' attitudes and readiness to implement self-evaluation (Chvál and Starý, 2008). An analysis of the content of self-evaluation activities developed by school principals revealed a clear need for external support; also in the absence of external specialist support there were serious drawbacks in learning exchanges among school principals in their peer review of each other's school development plans.

In the French Community of Belgium, school principal representatives report that developing a school plan can be a useful process if it valued and invested in by the whole school team (Blondin and Giot, 2011). If not, it is rather an imposed requirement that demands much work. Stakeholders also cite the need to lessen the high administrative burden on school principals in order to free up time for pedagogical leadership – this is particularly challenging in primary schools.

In Denmark, school principals have expressed concern about how to reconcile the increased external demands to document school quality with a strong climate of trust at their schools (EVA, 2007). This was echoed in school principal demands for training on how to meet documentation requirements (EVA, 2010). Another study concluded that 70% of school principals request professional development in evaluation, strategic development and quality assurance (Chairmanship of the School Council, 2009). National training sessions on working with evaluation are offered to school principals and teachers on a voluntary basis and in parallel online evaluation tools and examples are offered.

In France, there has been little if any emphasis on school self-evaluation in leadership training (Dos Santos and Rakocevic, 2012). As such, both the School of National Education, Higher Education and Research (ESEN) and the national education authorities at regional level (*autorités académiques*) have developed targeted training programmes, with a particular emphasis on how to use indicators. The Directorate of Evaluation, Forecasting and Performance (DEPP) provides packages to help schools build their own indicator sets. A broad cross-section of stakeholders report that school principals and teachers need training in how to conduct self-evaluations, including setting objectives and using the indicators and tools available to them.

An evaluation of the new approach to external school evaluation with a focus on school self-evaluation in Hong Kong-China identified self-confident and calm leadership as an important factor in helping embed a culture of reflection and inquiry (MacBeath, 2008).

In the Atlantic Provinces of Canada, the emphasis on school self-evaluation as a part of school improvement planning places high demands on school principals and other members of the school leadership (Fournier and Mildon, forthcoming). There are similarly high demands on school principals in New Zealand where the approach to school evaluation emphasises the importance of school self-evaluation (see Box 6.13).

Box 6.13 Targeted training on school self-evaluation for school principals in New Zealand

As school self-evaluation (self-review) is at the heart of the New Zealand approach, school capacity to conduct self-evaluation is of key importance. There are high expectations on school principals and their organising bodies (Boards of Trustees) and this can be especially challenging for schools in isolated areas or in communities with low socio-economic status. The external school evaluation body (the Education Review Office, ERO) in collaboration with school principal associations delivers workshops on self-evaluation to school principals, their teams and organising bodies and has developed support materials and case studies in good practice in self-evaluation. Such initiatives capitalised on ongoing professional development for external evaluations and so costs were minimal. In 2009, 35 workshops were delivered by a national facilitator and supporting local senior evaluators to over 1 200 participants across New Zealand, including relatively isolated areas. Workshop feedback was positive and external evaluators are reporting improved self-evaluation processes from schools that attended the workshops.

Source: New Zealand Ministry of Education (2010).

Creating evaluation roles and responsibilities within the school

Research from different countries has pointed to how the creation of teams holding particular responsibilities for self-evaluation within a school can positively impact the effectiveness of self-evaluation. In the Netherlands, the use of "data teams" comprising teachers and school principals or deputy principals has proved effective in encouraging greater use of data in school self-evaluation for improvement. This is a mechanism to engage school leadership support and to foster collaboration by focusing on specific problems in the school (Schildkamp, Rekers-Mombarg and Harms, 2012). In Estonia, although the school leader is responsible for implementing self-evaluation, many schools have developed "assessment teams" and there has been positive feedback on this (European Commission, 2011). In England in the United Kingdom, there is a "data

hierarchy" in schools with well-developed data and evaluation roles among school leadership and senior staff (Kelly and Downey, 2011). In Hong Kong-China a new focus on the role of self-evaluation within the external evaluation process has seen the creation of School Improvement Teams in schools. An evaluation identified that schools with effective and credible School Improvement Teams were able to cope better with the new external evaluation approach. The most effective teams comprised a cross-section of staff with high credibility among their colleagues, showing vision of how self-evaluation could feed into learning and school improvement, working as a team with distributed leadership, exercising initiative and creativity, and able to instil an ethos of accountability (MacBeath, 2008).

Engaging the full school community in school self-evaluation activities

School self-evaluation is seen as a way to engage all school staff in collective learning (Hopkins, 1995) and to aid the constant search for improvement (Barber, 1996). It is also important to engage key stakeholders in self-evaluation and improvement activities (see Box 6.10). Leithwood and Aitken (1995, p. 40) define a "learning organisation" as: "A group of people pursuing common purposes (and individual purposes as well) with a collective commitment to regularly weighing the value of those purposes, modifying them when that makes sense, and continuously developing more effective and efficient ways of accomplishing those purposes".

Importantly, research has underlined the important role that students can play in school self-evaluation. Evidence from several systems has highlighted that involving students in decisions about their schooling is an important factor in school improvement. Students have a critical role to play in determining how schools and classrooms can be improved (Rudduck, 2007; Smyth, 2007; MacBeath et al., 2000), even though they need support to learn how to provide powerful feedback (Pekrul and Levin, 2007).

Seashore Louis et al. (2010) conducted a six-year study in nine states, 43 school districts and 180 schools in the United States and found that higher-performing schools generally asked for more input and engagement from a wider variety of stakeholders. In all schools studied, school principals and district leaders had the most influence, but a greater degree of influence from teacher teams, parents and students did not lessen school principal influence.

Emerging evidence from a research project on the impact of external school evaluation indicates that schools reporting greater capacity for improvement and change, also report greater stakeholder involvement and greater efforts to improve teacher co-operation and school leadership (Ehren et al., 2012).

In the Netherlands, analysis of the use of self-evaluation in primary schools revealed that schools undertaking this as part of a "learning organisation" perform significantly better in mathematics (Hofman, Dukstra and Hofman, 2009).

The benefits of establishing a "professional learning community" are widely recognised. However, there are common challenges in many systems to instil this cultural change in many schools and also among staff within a school (see Box 6.14).

In Japan, there are official requirements for schools to seek the views and demands of students, mainly by means of questionnaires, in particular seeking their evaluation of teaching. The increased use of student evaluation of teaching and lessons in secondary schools had sometimes caused defensiveness and hostility among teachers (Katsuno and Takei, 2008). Although, there are no official requirements for student participation in

school self-evaluation forums (including parents, local residents, school counsellors), this had been encouraged by the Saitama Prefecture. During a study of self-evaluation in six schools during 2004-06, Katsuno and Takei (2008) comment that the Saitama Prefecture takes a less managerial approach with a greater emphasis on discussion and communication (p. 176). The study found some positive results: although it was not always an easy process to involve students (some teachers were not willing to listen to student demands), many teachers saw this self-evaluation process as a way to promote students' personal and social development.

In Slovenia, there is a legal requirement for the teacher assembly, the parent council and the school board to discuss the school's annual report, to evaluate the results and to include their evaluation, comments and proposals in the final report (Brejc, Sardoč and Zupanc, 2011). In turn, this evaluation process feeds into the school's development plan and annual work plan.

Box 6.14 Building professional learning communities: A challenge shared

Research has identified the school self-evaluation process as a matter for school leadership in many schools, with limited engagement from teachers in the process. Schools with highly developed self-evaluation processes promote a "professional learning community" in which each member of the school is constantly learning and improving.

Evidence from the OECD's Teaching and Learning International Survey (TALIS) of lower secondary school teachers in 2007-08 indicates that it is a common challenge to build professional learning communities in schools. TALIS provides measures on five different aspects which together would make a professional learning community: "co-operation" (Exchange teaching materials with colleagues; and Teach jointly as a team in the same class); "shared vision" (Attend staff meetings to discuss the vision and mission of the school); "a focus on learning" (Ensure common standards in evaluations for assessing student progress); "reflective inquiry" (Take part in professional learning activities, e.g. team supervision); and "de-privatisation of practice" (Observe other teachers' classes and provide feedback).

In all countries, there is great variation among teachers in the extent to which they indicate their participation in the five types of activity associated with a professional learning community. In each country there are three or four main teacher profiles according to the type of activities they participate in and how often they do so. A study in the Flemish Community of Belgium in 96 schools with 2 716 respondents also revealed significant variation in reported attitudes toward self-evaluation activities within schools, more so than between schools – and more positive attitudes were associated with a more pronounced professional learning community culture (Vanhoof, Van Petegem and De Maeyer, 2009).

The TALIS results show that in many countries basic forms of co-operation among staff are common, but it is much less common for teachers to work together on core professional activities (participation in reflective inquiry and observing other teachers' classes and providing feedback). Teachers in larger schools more frequently reported that they observe other teachers' classes and provide feedback.

Important factors associated with increased levels of participation in professional learning community activities include receiving feedback and appraisal on teacher instruction and being involved in external professional development activities. Both factors indicate the important role that the observation of the teaching and learning process plays, including potentially via critical friendship peer observation activities among schools.

However, results also show that being more actively involved in a professional learning community can be time consuming. Vanhoof, VanPetegem and De Maeyer (2009) found that teachers reported self-evaluation activities to be time consuming and difficult to carry out. Indeed, evidence from the Atlantic provinces in Canada indicates that this is a common concern raised among teacher unions (Fournier and Mildon, forthcoming).

Sources: OECD (2012); Vanhoof et al. (2009); Fournier and Mildon (forthcoming).

In the French Community of Belgium, each school has a participation council (legal requirement since 2007) to ensure the rights for parents and students to give feedback to the school, however, teacher representatives report that the participation of parents varies from school to school and is particularly weak in schools in less advantaged socio-economic communities (Blondin and Giot, 2011). They argue that this can reinforce inequities among schools.

Stimulating and supporting peer review among schools

The OECD Review revealed incipient practices of schools undertaking peer evaluation activities in several countries. Box 6.15 presents an overview of emerging peer reviews in the Flemish Community of Belgium. Seeking external ideas and support, including from other schools, is a feature of effective professional learning communities (Bolam et al., 2005). There is considerable evidence, for example from Finland, Sweden and England in the United Kingdom, that school-school partnerships, clusters and networks can provide mechanisms for sharing effective leadership as well as effective practice in a way that contributes to raising the performance of the member schools (Pont et al., 2008). Within the United Kingdom, executive leadership across partner schools in England has proven to be a very effective mechanism for raising the performance of underachieving schools (Hill and Matthews, 2010). A particular power of learning networks between schools is the sense of moral purpose around making a difference for all children – learning on behalf of others as well as with and from others – as was the case in the Networked Learning Communities programme, a large-scale enquiry and development initiative involving 137 networks (1 500 schools) in England between 2002 and 2006 (Jackson and Temperley, 2007). However, official evaluation of secondary school participation in school networks indicates that it is a challenge to engage the participation of the academically stronger schools (National Audit Office, 2009).

Critical friends are trusted outsiders. Frequently, they are external advisors, but the benefit of colleagues in other schools playing this role is that they are fellow professionals who are equal. They have the potential to hone pedagogic peer evaluation skills and to create the impression that schools are no longer alone. A study to promote school self-evaluation in 27 primary schools in the Netherlands also involved visitation by critical friends (Blok, Sleegers and Karsten, 2008). Schools principals reported that the use of critical friends was cost effective, although time consuming, and there was almost unanimous agreement that it had contributed to the school's capacity to improve (p. 391).

However, paying attention to the challenges and facilitating conditions for professional learning networks is critical to their potential to enhance educational change and support improvement (Chapman and Hadfield, 2010). Trusting relationships are necessary for deep networking and can be fostered by the prior agreement among participating schools on a code of ethics to guide the peer evaluation process (Stoll et al., 2011). The context in which schools conduct self-evaluation determines to a considerable extent the nature of the support that a critical friend can offer (Swaffield and MacBeath, 2005). If school self-evaluation is voluntary for the purpose of school improvement, a critical friend's role can be varied and potentially highly creative. However, if a school self-evaluation is mandated and subject to external evaluation, a critical friend's role is more politicised and there are higher stakes (p. 249).

Box 6.15 Emerging peer reviews among schools in the Flemish Community of Belgium

The Ministry of Education and Training has stimulated collaboration among schools by its promotion of "school communities" (*scholengemeenschappen*). Schools in a similar geographical area join a school community on a voluntary basis. However, the Ministry of Education and Training provides incentives for schools to join a school community by offering extra resources (i.e. extra teaching time for primary and secondary schools). In the case of secondary schools, there are also some organisational advantages to joining a school community. These efforts have successfully stimulated further collaboration among schools and virtually all schools offering mainstream primary and secondary education belong to a school community. There are clearly defined responsibilities for schools and belonging to a school community "implies continuous evaluation and adjustment of school policies" and therefore effectively promotes school improvement (Ministry of Education and Training and the University of Antwerp, 2010).

Although these emerging collegial relationships are at relatively early stages of development, their emergence is a strength in that they are focusing on helping schools develop both their self-evaluation capacity and the potential for critical friendship. The OECD Review visit revealed an example of primary school principals collaborating with colleagues observing teachers in each other's schools and an inter-schools quality network between secondary school principals focusing on how to stimulate and improve the use of outcomes. Research points out that schools find peer visitation a useful learning experience (Flemish Ministry of Education and Training, 2010).

Further, examples of peer visitation include: a project by the umbrella organisation for Provincial education, in which participating schools commit to a code of ethics, visiting teams write a report on findings, strengths, weaknesses and recommendations, and the visited school decides how to address these recommendations; and visits and peer reviews among schools involved in similar innovation projects in the City of Antwerp.

Source: Shewbridge et al. (2011a).

Reporting and use of results

Results from the three major approaches to school evaluation can have both formative and summative uses. For example:

- **School development**: external school evaluation can lead to recommendations or instructions on particular aspects for individual schools to improve and can be used to identify and share best practice and innovative practice throughout the school system; self-evaluation results can feed into the development of a school improvement plan and professional development activities; both types of school evaluation can use comparative school performance measures to identify relative strengths to build on and weaknesses to be improved.
- **School accountability**: external school evaluation results for individual schools or groups of schools (e.g. local authority overview reports) may be published, results may lead to possible rewards (e.g. national or regional competition recognition; additional funding) or sanctions (e.g. being publicly named as an underperforming school with quality concerns; loss of national recognition or funding; school closure) or strengthened external supervision and/or support; school self-evaluation results may be reported to the school community to give account of the school's status and progress toward specific school goals; school self-evaluation results may form the basis of external school evaluation; comparative school measures may be reported to the public for general accountability as well as to aid families in choosing schools (as individual school performance reports or in national or regional performance tables).

An overview of accountability uses

Table 6.11 provides an overview of the use of school evaluation results and school performance measures for accountability. This also gives a sense of the relative influence that the results of external school evaluation, school self-evaluation and comparative measures of school performance can have in school accountability. Countries are arranged in descending order of the possible influence that external school evaluation has on the evaluation of school performance. In general, across countries external school evaluation carries more influence than school self-evaluation and comparative school performance measures in terms of evaluating school performance and school administration and also for informing decisions on possible school closures and financial rewards or sanctions.

It is clear that in the majority of systems where comparative measures of school performance are available (via either national examinations, national assessments or both), these exert a high to moderate degree of influence over the evaluation of school performance. In fact, Korea and Spain are the only countries where these do not influence evaluations of school performance. In Chile and Mexico, comparative measures of school performance are the only means to evaluate school performance. While comparative measures of school performance have a high or moderate influence in the evaluation of school administration in a few countries (the French Community of Belgium, Chile, Hungary, the Netherlands, Poland and England in the United Kingdom), they very rarely feed into decisions on financial rewards or sanctions or possible school closure.

However, in the Netherlands and England in the United Kingdom, school performance measures do have some influence in decisions on possible school closures as these are used within the external school evaluation process – this is also the case for Ireland where these have a low degree of influence. In the Netherlands, comparative measures of school performance are a key part of the external school evaluation process. Although schools are not readily closed down (there is a procedure involving a number of escalating steps), their results in principle have high impact on their potential closure. As detailed in Box 6.6, school performance is a crucial indicator in the Netherlands for the Inspectorate of Education's decision making process to judge whether a school is "weak" or "very weak".

In Hungary there is strong reliance on the use of student assessment results in school evaluation. A survey of school organising bodies (maintainers) revealed that 84% reported relying on national assessment results and nearly half of local governments reported only using one source of information and this tends to be national assessment results. The publication of national assessment results "undoubtedly" qualifies as the initiative with the greatest impact on school evaluation activities and has been largely accepted (results of the Institute of Education Research and Development's 2009 school survey revealed that around 10% of school principals reported disagreement with public access to national assessment results). Schools failing to achieve a minimum level of performance in national assessments are required to take measures. Such legislation has been strongly criticised by the Education and Opportunities for Children Roundtable as a political idea that is not based on adequate evaluation (Kertesi, 2008, p.185, in Hungarian Ministry of Education and Culture, 2010).

Table 6.11 Use of school evaluation results and school performance measures for accountability (2009)

	Degree of influence the results of external school evaluation have over:				Degree of influence the results of school self-evaluation have over:				Degree of influence the results of national assessments or examinations have over:			
	Evaluation of school performance	Evaluation of school administration	Possible financial reward/ sanction	Possible school closure	Evaluation of school performance	Evaluation of school administration	Possible financial reward/ sanction	Possible school closure	Evaluation of school performance	Evaluation of school administration	Possible financial reward/ sanction	Possible school closure
Belgium (Fl.)[1]	High	Med.	High	High	a	a	a	a	High	Low	None	None
Belgium (Fr.)	High	High	None	Low	a	a	a	a	High	High	None	None
UK (England)	High	High	Low	High	Low	Low	None	Low	High	High	Low	Med.
Iceland[1]	High	High	None	a	High	High	None	a	High	None	None	a
Ireland	High	High	None	Med.	a	a	a	a	High	None	None	Low
Netherlands	High	Med.	None	High	a	a	a	a	High	Med.	None	High
Poland	High	Med.	m	None	Med.	Med.	m	None	High	Med.	Med.	None
Portugal	High	Low	None	None	Med.	Low	None	None	Med.	None	None	None
UK (Scotland)[2]	High	High	a	High	High	High	a	High	High	None	None	None
Slovak Republic	High	Med.	Med.	High	High	High	Low	Med.	Med.	Low	None	None
Turkey	High	High	Low	High	Med.	Med.	None	None	High	Low	None	None
Czech Republic	Med.	High	High	High	Med.	Med.	Med.	Low	a	a	a	a
France	Med.	Med.	Low	None	Low	Low	None	None	High	None	None	None
Austria	Med.	Med.	Med.	Med.	Low	Low	Low	Low	a	a	a	a
Germany	Med.	Med.	None	None	Med.	Med.	Low	None	High	Low	Low	None
Israel	Med.	None	None	None	Med.	None	None	None	Med.	Low	None	None
Korea[1]	Low	High	Low	None	High	High	Low	a	None	Low	None	None
Luxembourg	Low	Low	a	a	Med.	Low	a	a	High	Low	None	None
Spain[1]	Low	Low	None	None	a	a	a	a	None	None	None	None
Chile[1]	None	None	High	High	a	a	a	a	Med.	Med.	None	None
Estonia	None	None	None	None	Med.	High	Low	None	Med.	None	Low	None
Hungary	a	a	a	a	m	High	None	None	Med.	High	None	Low
Mexico[1]	a	a	a	a	a	a	a	a	Med.	None	None	None
United States[2]	m	m	m	m	m	m	m	m	High	m	m	m

Notes: The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing.
(1) Comparative measures of school performance refer to national assessments only.
(2) Comparative measures of school performance refer to national examinations only.

Source: OECD (2011).

The results in school performance accountability systems may be used to identify underperforming schools and to target support to these schools. Evidence from the United States highlights the importance of providing adequate support to schools in need of improvement. The United States Department of Education commissioned an evaluation of school improvement support offered to schools identified in the No Child Left Behind accountability system (Padilla et al., 2006). The evaluation was conducted over three school years from 2001-04 with annual surveys to a nationally representative sample of 1 300 district administrators and 739 schools, plus case studies in 20 schools, in 15 districts across five states. By 2003/04 almost all districts provided identified schools

with basic school improvement support, e.g. writing an improvement plan and analysing data. However, all but two of the 15 districts in the case study reported serious capacity concerns to offer school improvement support, including reduced funding available for teacher professional development and a lack of knowledge and skills to provide school-based instructional support. The study also highlighted the dominance of contextual school characteristics influencing whether or not schools improved enough to exit "improvement status". The authors conclude that identified schools require much more intensive support.

Regarding the nature of support, a study of 21 low-performing high schools across six states in the United States revealed that school stakeholders appreciated the experience, dedication, interpersonal skills and accessibility of their support providers (Boyle et al., 2009). This shows appreciation for flexible and adapted support according to the school context and that such an approach is perceived by schools as high-quality support. Schools also noted the importance of the intensity, stability and timeliness of the support offered. This highlights the importance of the quality and capacity of the external support providers.

School self-evaluation

The results of self-evaluation are primarily aimed at making plans for school development and further professional development needs. For example, in Slovenia, the annual school report must be discussed with and evaluated by the Teachers' Assembly, the Parents' Council and the School Board. Each group can comment of the effects of the school's programmes and policies and make proposals on how to develop these. This internal evaluation subsequently feeds into the school's development plan and its specific annual work plan (Brejc, Sardoč and Zupanc, 2011). In New Zealand, schools are expected to integrate the results of both their own self-review and national external school reviews into their long-term planning (Nusche et al., 2012).

In many countries, school self-evaluation is integrated in a classic management approach to strategic improvement planning. For example, following Deming's Plan-Do-Check-Act cycle (based on Hofman, Dukstra and Hofman, 2005):

- At the plan stage schools may: conduct an analysis of strengths, weaknesses, opportunities and threats (SWOT); develop a vision/mission; set goals; develop specific, measurable, acceptable, realistic and time-bound (SMART) performance indicators; establish functional structural communication; create broad support for the plan; and plan funding and staff allocation.
- At the do stage schools implement the plan and may: ensure educational leadership; apply policy and strategy; work out an activity plan; stimulate a professional culture; ensure internal and external communication.
- At the check stage schools evaluate the plan and may: ensure internal and external involvement in evaluation; use monitoring; analyse data; construct performance indicators and norm references for evaluation; analyse staff and student satisfaction with the improvement plan; report results to the school community.
- At the act/adapt stage schools act on the evaluation results and adapt the plan accordingly: integrate findings in the monitoring system; broaden the application area; deploy necessary staff and material; check failures; seek accreditation; restart the cycle.

Box 6.16 Using school self-evaluation results in school improvement planning

Quality management in schools in Sweden

In Sweden, internal quality management in schools (stimulated by long established quality reporting practices) fosters the intelligent collaborative use of feedback. Also, the relatively intensive school self-evaluation activities contribute to the openness of professionals for feedback coming from external school evaluation. In Sweden, feedback seems to be integrated in schools into a communication-rich organisational environment which is capable to understand and interpret it.

The concept of quality management or quality development, as it is reflected in the quality model developed by the National Agency for Education, is embedded in a classic strategic management model focusing on four key questions: (1) who are we?, (2) where do we want to go?, (3) how can we get there?, (4) how did we succeed?. This is the complete strategic planning cycle which starts with a self-analysis and the analysis of the environment, it continues with vision-making and strategic goal setting, then implementation planning and, later on, the evaluation of the results. Quality reporting is, in fact, only the last element of this process, its most important aim being to feed back into the four-stage strategic cycle.

The typical approach to self-review within a school planning cycle in Australia

In Australia, school planning is a continuous process best understood as cyclical, developmental and adaptive. All state and territory schools are committed to self-reflection, strategic planning and transparency of reporting when evaluation and assessing their individual schools performance. School self review is the first step in the process of school development and improvement, providing the foundation for reporting and accountability. School self-assessment practices are performed in all public state and territory schools. This is through a process of monitoring and assessing yearly operation plans, strategic plans and measuring against key performance measurement indicators, as established by individual state and territory guidelines. Self review enables an analysis of current performance and the effectiveness of strategies implemented to support performance improvement. It provides the basis for performance reporting and future improvement planning. School self-assessment appears to be most effective when assisted by significant levels of support from the state and territory departments or school regulatory bodies, especially in the form of external reviews and the provision of templates and standard frameworks.

School self-review steps and procedures in Newfoundland and Labrador, Canada

For a number of reasons, Newfoundland and Labrador has been particularly successful implementing a School Improvement Program. Historically, they had schools working on models of improvement as early as 1986 and they did a pilot project and study in 1995 (Sheppard, 1995), adopting a model revised from that experimentation in 2004. However, senior department officials attribute the effective implementation to the support system and capacity-building put in place.

While there are many methods to gather, record, analyse, and make informed decisions, the steps below have been field-tested in schools and have been found to be effective. A timeline is also suggested for each of the steps. It is recommended that the Internal Review component be completed within a 5-month period, though this is sometimes contingent upon the nature and culture of the school.

- Step one: Establish a school development (leadership) team
- Step two: Gather and organise relevant data according to criteria statements
- Step three: Establish data recording and analysis teams
- Step four: Record and analyse the data
- Step five: Report on data and critical issues
- Step six: Goal identification

Sources: Nusche et al. (2011b); DEEWR (2010); Fournier and Mildon (forthcoming).

By design, this implies that the results of school self-evaluation feed into school development policies. Box 6.16 presents typical examples of the approach to self-review as part of school improvement planning cycles in Australia, Canada and Sweden. These approaches clearly illustrate that school self-evaluation is fundamentally integrated in a broader strategic planning cycle. For example, in the Atlantic Provinces of Canada the final step in the school improvement planning cycle is the external validation of the school's results. A school completing the strategic improvement planning cycle by definition has improved (Fournier and Mildon, forthcoming). The examples also illustrate that schools with effective school improvement planning establish clear procedures and can benefit from well-developed external support systems.

Schools may also choose to publish annual reports on their websites. However, some systems may set requirements for schools to publish annual reports. For example, in the Czech Republic schools have been obliged to publish their results and this is checked as part of the criteria in the inspection framework (Santiago et al., 2012b). As of 2012/13, there is a clear expectation that schools use the results of their self-evaluation activities as well as the educational results in publishing their annual school report and that this is all checked via external school evaluation by the Czech School Inspectorate.

Using comparative school performance measures in self-evaluation activities

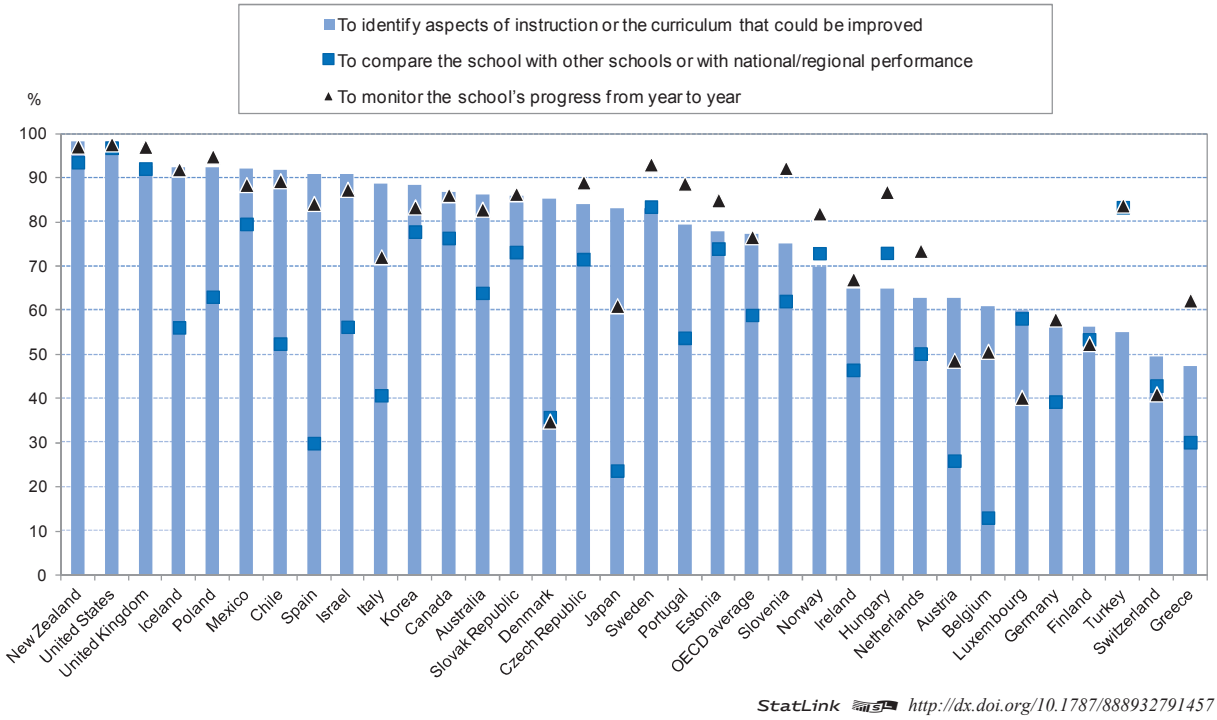
The OECD Review revealed in several countries that there is often a need to optimise feedback of results from both school-based tests and also national student assessments and surveys. When there is a lack of timely feedback or reaction to the results of internal school evaluations (quantitative or qualitative surveys) this lessens the support of educators for evidence-based self-evaluation. In a similar vein, typically, schools and teachers do not receive feedback reports from national assessments in time to diagnose the learning needs of the students tested. Such delays in feedback may lessen their use in school self-evaluation (and risk to lessen the support of educators for such national assessments who may perceive them as distant and of little value or use to them and their students). Finally, when schools receive feedback from central systems the results may well remain in the realm of school leadership and not be widely used or discussed among staff.

School performance feedback systems can be powerful tools providing timely, high-quality information on performance that the school can use for improvement actions (Visscher and Coe, 2003). This may help schools identify problems sooner and examine which types of interventions work better in different contexts. The availability of computerised systems for information processing has made a significant contribution to the logistics of school performance feedback (Visscher, Wild and Fund, 2001). Timely feedback of performance data in an accessible format are important characteristics of data systems that can promote the use of results in schools. However, of equal importance are characteristics related to the users and the school organisation (see Box 6.17). Research on stakeholder perceptions of the use of data in the United States revealed that untimely feedback of performance data coupled with a lack of resources to support data use, e.g. extra time, staff or training, may mean that data are irrelevant for teachers (Englert et al., 2007). Further, the research revealed much more positive attitudes toward data use in improving schools.

According to school principal reports during the PISA 2009 survey, the use of assessment data for school self-evaluation is pretty well established in secondary schools across the OECD. On average in the OECD, 77% of students are in schools whose principal reports the use of student assessment data to monitor the school's progress from

year to year and to identify aspects of instruction or the curriculum that could be improved (Figure 6.6). In particular, these results indicate that the use of assessment data is well established in New Zealand, the United Kingdom and the United States. In all but two countries, at least 50% of students are in schools whose principal reports that assessment data are used to improve the curriculum – and this is at least 80% of students in 18 countries. This indicates a strong use of results for development purposes. The results also indicate that in many countries assessment data are less often used for benchmarking purposes, that is, to compare the school's performance with other schools or with national or regional performance. Indeed, there appears to be a missed opportunity to feed student performance data from standardised tests into self-evaluation activities in some systems. Notably in Luxembourg, Finland and Denmark, although the majority of students are in schools reporting the use of standardised tests, there is little reported use of assessment data for school monitoring (Figure 6.5).

Figure 6.6 Use of student assessment results in school self-evaluation (PISA 2009)



StatLink http://dx.doi.org/10.1787/888932791457

Note: Percentage of students in schools where the principal reported assessments of students in the national modal grade for 15-year-olds are used for these purposes. Data are shown for OECD countries.

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)* , OECD Publishing, Paris.

Of course, such self-reports do not shed light on the actual quality of use of student performance data. In Ireland, one of the findings in primary schools from the 2009 national assessments was that while the aggregated results of students' standardised test results were widely discussed at staff meetings, the use of such data to establish school-level learning targets was less common (Eivers et al., 2010). This is backed up by external school evaluations revealing limited capacity for schools to monitor progress (Irish Department for Education and Skills, 2012). In New Zealand, the Education Review Office identified via

external school evaluation in 2007 that only 17% of schools used student achievement data to aid decisions on meeting learning needs of nationally identified priority student groups (ERO, 2007). A review three years later identified that although two-thirds of schools used assessment information to identify "at risk" student groups, only some schools used this to identify talented students who may require extra challenge (ERO, 2010). Some schools took actions based on the data to better meet the needs of identified students, however, few schools reviewed the effectiveness of these actions.

In the United Kingdom, research in 178 secondary schools in England sheds light on how the external school evaluation body's online self-evaluation software is used (RAISE online) (Kelly and Downey, 2011). School principals receive access codes and an administrator account and can control access to the software. 95% of teachers in the survey responded that they did not have access to RAISE online. This reflects the typical approach of data management responsibilities lying with senior managers and the feedback to teachers of pre-analysed data – presumably to increase efficiency and provide teachers with "information" rather than raw data. The research revealed that teachers would prefer to analyse data in teams or within departments.

In New Zealand, the United Kingdom and the United States, both the publication of achievement data and the use of external benchmarking in self-evaluation activities are reported widely (see Figures 6.6 and 6.7). School principal reports indicate that the tracking of achievement over time by administrative authorities has an influence over their use by schools (see Chapter 8).

Box 6.17 Factors promoting use of assessment data for self-evaluation

The literature identifies three major factors promoting data use in schools:

- Data system characteristics – timeliness of data availability, accuracy, validity, relevance and reliability of data; access to data; tools available to use the data.
- Data user characteristics – whether they believe in the data, have the necessary knowledge, skills and motivation to use them, whether they feel power to make changes (or whether they feel improvement is contingent upon things beyond their control).
- School organisational characteristics – time is allocated for data use, colleague(s) with special role/expertise in data use, training, teacher collaboration, data use is linked to school vision, norms and goals, school principal supports data use.

Source: Schildkamp, Rekers-Mombarg and Harms (2012).

In Austria new annual assessments against national standards have been introduced in 2011/12 and include feedback reports to schools showing school and class level aggregate results with the aim of promoting their use for school development activities (Specht and Sobanski, 2012). This reporting system is also accompanied by moderators trained by the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) who are made available to schools to aid their interpretation of the results. During the piloting of these new assessments in 2009 in 204 secondary schools, this feedback process was evaluated via surveys to teachers and school principals. Results indicate that such feedback is useful for schools with school principals reporting that this had stimulated professional communication at their school, including 40% reporting that results had already led to decisions to make changes at the school and 30% of teachers reporting that results might have an impact on their instruction (Grillitsch, 2010). There are

also plans to introduce experts to support schools with their quality development. Of note also, 40% of teachers reported being encouraged by the feedback and only 3% discouraged.

In Norway, the leaders of regional education authorities see untapped potential in examination results, although some school principals do use these for self-evaluation activities (Roald, 2010). Schools with established self-evaluation cultures, the ability to interpret results and to design and implement measures according to the results benefit most from national assessment and examination results (Roald, 2010; Langfeldt et al., 2008). An evaluation of the national assessment system in 2009 revealed that around 50% of school organising bodies, school principals and teachers reported that they followed up on the results from the national pupil survey and that normally the results are discussed with teacher teams within the school (Allerup et al., 2009). Further, there is a strong relationship between a school's constructive use of national assessment results and the school organising body's capacity to support schools in their self-evaluation. The information provided by the national assessment system can seem to create bureaucratic work and take too much time if the information is not viewed as relevant and analysed for school development. Finally, there is a legal requirement for school organising bodies to follow up on the results of parent surveys in an attempt to engage parents in school development discussions. However, the OECD Review identified that data in the national School Portal was not extensively used by schools – when schools did use this, it was not at a whole-school level, but rather at a teacher level (Nusche et al., 2011a). This may be due to a lack of sophistication in the data presentation for analysis, for example, no inclusion of school contextual information to make more meaningful comparison, but also to do with a lack of data-handling skills among school principals and their staff.

In Luxembourg, national assessments have been recently introduced. While stakeholders agree on the importance of using evidence and data for school improvement, the results of national assessments are not yet perceived to add value for improving teaching and learning in the class and so are regarded as taking time and limited resources for little value (ADQS, 2011). This is linked to an initial lack of national capacity to sufficiently exploit much nationally collected data and to respond to the demands of schools. There were concerns on the timeliness of feedback and the level of feedback. For example, school principals initially only received aggregate performance distribution for the whole school per subject, which limits the analytical value of the results for school improvement. In addition, the introduction of national assessments has met with social resistance from teachers for fear of causal inferences being drawn from student assessment data.

In the French Community of Belgium the overriding impression from stakeholders is that there is too much information, but that it could be useful if there were clear guidance on how it could be used (Blondin and Giot, 2011). Teacher unions report that this equates to additional and meaningless administrative burden; school principals would like clear indicators for a quick overview of the major points and to answer key questions; parents underline the necessity of having clear explanations accompanying the results. School organising bodies see strong potential in the results, as long as these are discussed with a view to supporting and not controlling schools, which underlines the value of pedagogical advisors (a support service set up in 2007 for schools identified by external school evaluation or the school organising body as needing assistance).

In Canada, teacher federation position papers show that generally teachers in the Atlantic Provinces are still not convinced of the value of large-scale assessments (Fournier and Mildon, forthcoming). However, the results of large-scale assessments are included in school improvement plans and annual reports, which indicates that teachers

are using the results to monitor student performance and to evaluate the success of their long-term objectives. There are other indications from teacher federation position papers that accountability-related tasks weigh heavily on teachers' time and detract their core work with students.

External school evaluation

Table 6.12 presents more detailed information on the possible use of the results of external school evaluation for accountability. The columns to the left show where the influence of external school evaluation is greatest and the columns to the right show where it is weakest. In general, external school evaluation results are not strongly linked to financial rewards or sanctions. External school evaluation results do not impact the size of the school budget or teacher pay and bonuses (only a moderate influence in Austria, the Czech Republic and England in the United Kingdom), but do have a high degree of influence on other financial rewards or sanctions in the Flemish Community of Belgium and the Czech Republic. There are very different policies across countries on the degree of influence that external school evaluation can have on the possible closure of a school. In Estonia, France, Germany, Israel, Korea, Poland, Portugal and Spain the results of external school evaluation have no influence on possible school closures. However, external school evaluations have a high or moderate influence on possible school closures in nine OECD countries.

Table 6.12 Use of the results of external school evaluation for accountability (2009)

	The degree of influence the results of external school evaluation may have over:							
Total number of systems by level of influence:	Evaluation of school performance	Evaluation of school administration	Evaluation of individual teachers	Support to improve teaching skills	The likelihood of a school closure	Another financial reward or sanction	The size of the school budget	Teacher pay and bonuses
High	11	8	6	5	7	2	0	0
Moderate	5	7	8	9	2	2	2	3
Low	3	3	3	5	1	4	7	0
None	1	2	2	1	8	9	10	12
UK (England)	High	High	High	High	High	Low	Low	Moderate
Ireland	High	High	High	High	Moderate	None	None	None
Belgium (Fr.)	High	High	High	High	Low	None	None	None
Turkey	High	High	High	Moderate	High	Low	Low	None
Iceland	High	High	Low	Low	a	None	None	a
UK (Scotland)	High	High	a	High	High	a	a	a
Poland	High	Moderate	High	High	None	m	None	m
Slovak Republic	High	Moderate	High	Low	High	Moderate	Low	None
Netherlands	High	Moderate	Moderate	Moderate	High	None	None	None
Belgium (Fl.)	High	Moderate	Moderate	Moderate	High	High	Low	a
Portugal	High	Low	Moderate	Moderate	None	None	None	None
Austria	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Germany	Moderate	Moderate	Moderate	Moderate	None	None	None	None
France	Moderate	Moderate	Moderate	Low	None	Low	Low	None
Israel	Moderate	None	Low	Moderate	None	None	Low	None
Czech Republic	Moderate	High	Moderate	Low	High	High	Moderate	Moderate
Korea	Low	High	None	Moderate	None	Low	None	None
Spain	Low	Low	Low	Moderate	None	None	None	None
Luxembourg	Low	Low	Moderate	Low	a	a	Low	a
Estonia	None	None	None	None	None	None	None	None

Note: The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing.

Source: OECD (2011).

Mechanisms to ensure that schools follow up on the results of external school evaluation

There is a clear assumption that schools will use the feedback from school external evaluation and implement policies to address any weaknesses identified (e.g. Faubert, 2009; Ehren et al., 2012). However, evidence indicates that not all schools do this and that accepting feedback does not necessarily lead to school improvement actions. A degree of external follow-up can ensure that schools use external evaluation results to undertake school improvement actions (see section on Impact above). However, providing adequate follow-up can place significant demand on the external school evaluation body's capacity. Several countries take a policy to more closely supervise underperforming schools by the school inspectorate or review body and less frequent and/or less extensive review of well-performing schools (e.g. the Flemish Community of Belgium, Korea, the Netherlands and New Zealand). Box 6.18 presents an example of a strengthened follow-up of the results of external school evaluation in Ireland.

As of 2011/12, Portugal has also implemented a requirement for each externally evaluated school to prepare an improvement plan to respond to the challenges identified in the external school evaluation (Santiago et al., 2012a). The expectation is that each externally evaluated school will be followed up by educational authorities to assess the extent to which its improvement plan is effectively overcoming the shortcomings identified in external school evaluation. In the previous external school evaluation cycle, there was a lack of clear follow-up by the external school evaluation body, except in the most critical cases. A more systematic follow-up of schools may help to increase the impact of external school evaluations, as the OECD Review had identified that findings were not widely known among school staff. In Korea, as well as the closer follow-up of underperforming schools, schools receiving excellent evaluations are provided with level-differentiated support (although this has been evaluated as having limited impact, Jung et al., 2008).

There may also be specific expectations for the school to follow up on the results of external evaluation as an internal matter. For example, in the Flemish Community of Belgium there is an obligation for school leadership to discuss with school staff the results in an external school evaluation report on its school within 30 days (Flemish Ministry of Education and Training, 2010).

In Denmark the quality reports compiled by local authorities on their schools are not connected with any prescribed set of rewards or sanctions, but each local authority can set a specific action plan for schools with poor performance (Danish Ministry of Education and Rambøll, 2011). In their third year of implementation, the Danish Evaluation Institute judged that the form of the quality reports was adhered to, but that local authorities were struggling with how to follow up underperforming schools and to provide the required support (EVA, 2009). While around 73% of local authorities initiated follow-up activities in the 2006/07 school year, these were typically at the school system level and not for individual schools.

Box 6.18 Follow up on implementation of external school evaluation recommendations: Ireland

A follow-up mechanism is now in place in relation to schools that are identified during inspections as experiencing significant difficulty and where it is evident that intervention is required to assist the school improvement agenda. The School Improvement Group (SIG), was established by the Department in 2008 to ensure that improvement happens following inspection. The SIG, which comprises members of the Inspectorate and officials from the Department's Schools Division and other relevant divisions, co-ordinates the Department's actions in following up on the recommendations from inspections. The actions co-ordinated by this group are tailored to the specific needs of the school, and are intended to ensure that the school's patron, management and staff work to improve the quality of provision for students. Follow up on the implementation of recommendations in inspection reports comprises a number of differentiated approaches depending on the challenges facing particular schools. The range of interventions used to promote action and improvement include:

- meetings with the school patrons/trustees, chairpersons of boards and/or school principals
- progress reports from the board of management
- support for the school from school support services or services provided by patron or management bodies
- further inspections
- sanctioning school management, where warranted.

The School Improvement Group has been successful in helping an increasing number of schools to improve. Between 2008 and 2011 it has dealt with more than 50 poorly performing primary and secondary schools. In the majority of cases, the SIG has requested the boards of management of the schools to provide detailed progress reports on improvement. Meetings have been held with boards of management and school patrons and have resulted in a number of significant actions to put improved school governance arrangements in place. In some instances, school principals or other members of staff have resigned. School support services have also provided additional guidance and professional support to some schools. Progress is monitored through seeking update reports from the schools' management and by means of further inspection activity. The outcomes are reviewed carefully by the School Improvement Group. In common with initiatives to improve seriously under-performing schools in other countries, the experience of the School Improvement Group has shown that it can take some time to achieve significant improvement. However, approximately one-third of schools that came to the attention of the School Improvement Group are no longer in the process and there is evidence of significant improvement in a further third. The remainder have either entered the process recently or continue to be a cause for concern.

In 2012 "follow-through inspections" were trialled in a sample of schools that had been inspected over the previous three years. Such inspections aim to determine the extent to which the school has made efforts to improve practice and the progress that the school has made on implementing recommendations made during the previous inspection. The intention is to mainstream follow-through inspections and to publish associated reports.

Source: Irish Department for Education and Skills (2012).

Publishing results of external evaluation for individual schools

It is common practice in OECD countries to publish the results of external school evaluation in the form of a summary report on major findings within the school system (see Chapter 8). However, there has been increased demand to also publish external evaluation findings for individual schools, so that parents can use evaluation results in making decisions on which school their child should attend and also to more regularly follow quality developments at their school. The routine publication of external evaluation findings on individual schools has become increasingly widespread in Europe

and now happens in 16 countries or regions of Europe (Eurydice, 2012). However, the format that results publication takes varies significantly among countries, ranging from full evaluation reports with specific summaries, to just a few headline points on the major findings. The publication of individual school evaluation results has led to the need to communicate in a way that is more accessible to parents and the wider public in general (see Box 6.19).

In Korea, 2011 saw the inclusion of school evaluation results in the School Information Disclosure System, so that comprehensive school evaluation results on the school's strengths and recommendations were made publicly available for the first time. This is expected to give more weight to the impact of school evaluation (Kim et al., 2010). However, before this date, the highest and lowest performing schools were indirectly revealed via sanction and reward arrangements.

The decision to publish external school evaluation reports in the Flemish Community of Belgium aimed to meet parents' rights to clear and accurate information to inform school choice. In doing so, the Flemish Inspectorate of Education performed a balancing act: to provide relevant and useful descriptive information on school performance, but not performance measures that could be used to rank schools; to ensure that schools would not doctor or limit their self-evaluations for concern that key findings would be made available to the public via external school evaluation reports (Flemish Ministry of Education and Training, 2010). The decision to publish external school evaluation reports has led to external school evaluators using a more coherent format of reporting. Research also shows that different stakeholders believe that publication will lead to better self-evaluations by schools, better parental involvement in the school, including following school development (Vanhoof and Van Petegem, 2005).

Similarly, in Ireland the publication of external school evaluation reports (since 2006) does not include numerical data that could be used to compile league tables, although such data are extensively used during school external evaluation (Irish Department for Education and Skills, 2012). There is legal provision that the Department for Education and Skills can refuse access to information that can be used to compare the academic performance of schools (Education Act, 1998).

Box 6.19 Reporting results of external school evaluation

In the Netherlands, much effort has been invested in producing quality information that can be used by parents and schools together. Much of the Inspectorate's information on secondary schools is provided in a special website (Windows for Accountability). The information does not refer to inspection criteria or norms (so for example, for the broader public, no school is judged as being "very weak" on indicators such as student achievement results or student satisfaction). Schools also can access additional information via a private login and can see their relative performance (e.g. with percentile scores) on different indicators. The information base is partly the same as used by the Inspectorate. Schools also can put information on this website that has a benchmarking goal. The visual presentation is very attractive for the broader public.

In England within the United Kingdom, the inspectorate (Ofsted) presents inspection reports on line for each school. Each report includes: a brief summary of the key findings and major recommendations for school improvement; a glossary of key terminology used in the report to make this more accessible to the school community; and a letter to students thanking them for their co-operation and explaining in a clear and simple manner the main findings and recommendations.

In New Zealand, the Education Review Office publishes individual school review reports on its website. Although these reports are not intended for comparing or ranking schools, the media may attempt to rate or rank schools on the findings – in particular, review reports for secondary schools include examination results and value-added indicators. Due to the publication of review reports, schools perceive these as high-stakes evaluations (New Zealand Ministry of Education, 2010).

Reporting comparative school performance measures

The OECD Review has revealed varied policies regarding the reporting of comparative school performance measures. Some systems may report national examination results at the school level, but decide not to publish the results of national assessments at the school level, preferring these to be used exclusively by schools for school development. There are also wide differences in the sophistication of how results are reported, from the publication of static performance tables to the provision of interactive websites where parents and other stakeholders can select and compare information on different schools. Importantly, there are also differences in what is reported: whether this is simply the raw results of student assessments or examinations aggregated to the school level; or whether this is accompanied by contextual information on the school and general information to help parents and other stakeholders interpret the results. Box 6.20 provides an overview of some approaches to reporting comparative performance measures.

In Mexico, there is a new database of information (RNAME) providing useful information on each school, including quantitative data on student outcomes in the national assessments (ENLACE). These arrangements represent a good step forward in providing parents, local communities, educationists and the general public with some key information about schools both globally and individually (Santiago et al., 2012). The inclusion of school-level data on students' results in ENLACE assessments over a three-year period is a good feature. However, there is no information available on the qualitative aspects of school work, which lessens the use of the information for parents. As yet, no data are provided on the context of the school and it is not possible to compare similar schools.

In the Slovak Republic, static comparative data tables showing school average results in the national assessments are published on the national testing institute (NÚCEM) website (Shewbridge et al., forthcoming). Efforts have been made to improve the presentation of performance data to allow comparison of results in different regions of the Slovak Republic. A non-governmental organisation has recently developed a school performance website. This aims to present information from the national assessments and also information gathered from external school evaluations, but procedures to validate the information presented are not clear.

Schools may report comparative school performance information to parents directly. School principal reports in PISA 2009 indicate varied policies in this respect for schools attended by 15-year-old students both within and among countries (see Figure 6.7). For example, 50% of 15-year-old students or less in all but two countries (the United States and Turkey) are in schools where parents receive information to compare their child's school group performance with that of students in other schools. This could imply the availability of comparable school-level performance measures to only some schools within the country or large variation in how these are reported to parents among schools within each country. It is more common practice among countries for schools attended by

15-year-old students to report student performance relative to national or regional benchmarks to parents, but again this varies considerably among countries (above 70% of students in 11 countries and below 30% in 10 countries).

On average in the OECD, less than 40% of 15-year-old students are in schools whose principal reports that achievement data are posted publicly (Figure 6.7). This is rare practice in Austria, Finland, Germany, Ireland, Japan, Spain and Switzerland; but much more typical in New Zealand, the United Kingdom and the United States. In the French Community of Belgium, schools cannot communicate student achievement data for promotional or competitive purposes.

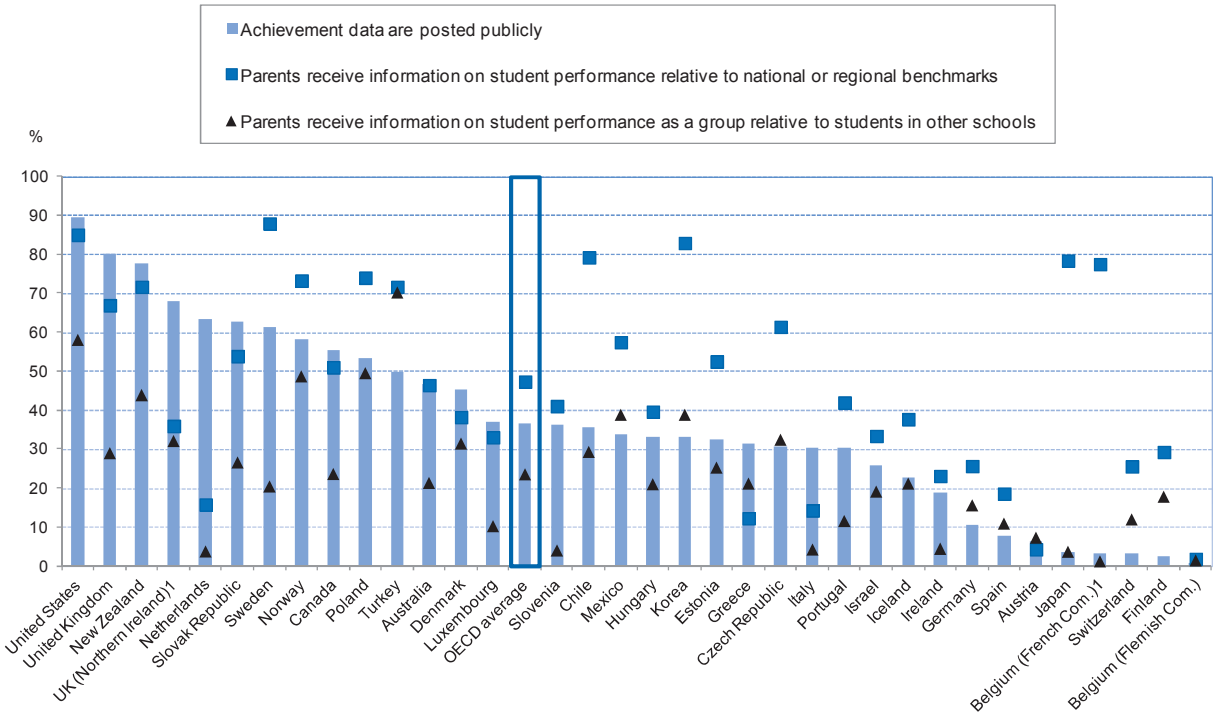
The publication of school performance measures aims to stimulate teacher and student efforts to improve performance, provide information to parents for school choice and stimulate improvement through competition, and reduce asymmetry of information providing a basis for more effective allocation of resources (Faubert, 2009). In an overview of literature mainly from the United States, the United Kingdom, the Netherlands and Norway, Rosenkvist (2010) finds that the available evidence regarding the effect of publishing student test results in school performance tables is mixed. There is wide consensus in the literature that reporting student test results in performance tables is coupled with several methodological problems and challenges. Notably, school performance tables are only really able to meaningfully distinguish schools performing at the top and bottom ends of the performance distribution, but performance differences between the majority of schools are rarely significant. There is little evidence of a positive relationship between performance tables and increased student performance. There is, however, evidence of performance tables influencing the behaviour of schools, teachers and parents – although not always as originally intended by the authorities. Research in England in the United Kingdom identifies a high degree of stress on English and mathematics departments due to the importance of these subjects in school performance tables, but that these departments had a sense of more power within the school, with easy access to school leadership and strong cases for negotiating extra resources and curriculum time (Perryman et al., 2011).

Studies have shown that school agents may view the publication of student achievement data as carrying high stakes even when the results are used only to identify areas for school development and are not linked to rewards or sanctions. Consequently, school agents such as teachers will work to avoid the public stigma of poor results, and this may have unintended consequences on classroom teaching and assessment (Corbett and Wilson, 1991; Madaus, 1988; McDonnell and Choisser, 1997). Such unintended consequences may include curriculum narrowing, teaching to the test and emphasising basic knowledge and skills that are easily measurable. Teachers may be tempted to design their own assessment in similar ways to the national assessments (i.e. typically in multiple-choice and short answer formats) to the detriment of richer, more performance-based approaches to assessment (see Chapter 4).

However, the fact that school leaders and teachers respond strategically to national assessments or national examinations implies that these can be a powerful tool to steer what is taught in classrooms. Advocates for the use of assessments in school performance accountability systems argue that teaching to the test content is appropriate if tests are properly constructed to measure achievement (Sims, 2008). Rosenkvist (2010) presents an overview of studies showing that the publication of school performance measures is associated with teachers emphasising the content that is prioritised in explicit policy goals, e.g. raising basic skills. This of course heightens the

importance of appropriately designed assessments. Other significant concerns raised in the literature regard a gaming and outright cheating culture in schools that can arise in response to incentives in school performance accountability systems. All of these concerns underline the importance of well-designed assessments that reduce the predictability of the assessment and the susceptibility of the tasks to inappropriate test preparation (Koretz, 2010).

Figure 6.7 Public reporting of student performance and reporting to parents (PISA 2009)



StatLink http://dx.doi.org/10.1787/888932791476

Notes: Percentage of 15 year-old students in schools where the principal reported that student performance information was reported in this way. Data are shown for OECD countries.

(1) Caution is needed when comparing results, as these were not internationally adjudicated.

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)* , OECD Publishing, Paris.

The publication of performance tables based solely on "raw" student test results essentially measure the quality of the school intake rather than the teaching in the school (Willms, 1997; Hoyle and Robinson, 2003 in Rosenkvist, 2010). One factor here is that results in national assessments and examinations and socio-economic status are highly correlated, such an approach risks to measure in part the socioeconomic characteristics of the students in the school rather than the contribution of the school to student learning (Faubert, 2009). This may lead to misinterpretation of school performance rankings that are damaging to equity and create incentives for schools not to accept students from disadvantaged socio-economic background (idem.). Another factor is that performance tables do not take into account students' prior achievement levels. In turn, this may create incentives for schools not to accept students with a less academic profile or special

educational needs. For these reasons, the OECD (2008) argues that although the publication of actual student assessment or examination results provides some important information, these are "poor measures of school performance".

There are different ways to address the concern about differences in school contextual characteristics. In the simplest form, school performance measures may present actual student assessment or examination results, plus provide descriptive information on the school context, for example school-level information on the students that participated in the assessment or examination, e.g. their gender, socio-economic background, etc. Statistical adjustments may also be used to account for the school context's impact upon a specific set of student assessment or examination results (these are referred to a contextual attainment models) (OECD, 2008). However, these do not take into consideration a student's prior attainment.

A substantial improvement in the presentation of school performance information is the use of statistical models that aim to measure the "value added" of a school. These are defined as "a class of statistical models that estimate the contributions of schools to student progress in stated or prescribed education objectives (e.g. cognitive achievement) measured at least at two points in time" (OECD, 2008). Therefore, a measure of value added would show the progress in student learning at a given school by taking into consideration a student's prior attainment. Further, these could also adjust for the school's context (contextual value added models).

However, research has also identified some concerns over the use of statistical models to adjust school performance measures (Rosenkvist, 2010). In a review of different statistical approaches for school performance measures, Masters (2012) concludes that these will inevitably provide an imperfect picture of a school's effectiveness. One concern in more complicated statistical adjustment models is a level of obscurity that makes it difficult for users to meaningfully interpret results. Such complex models may be open to accusation of massaging the results to make these appear better (e.g., van de Grift, 2009) or even to excuse low average performance (Figlio and Loeb, 2011) or to institutionalise low expectations (Hamilton and Koretz, 2002).

There is evidence that educators support the use of school performance measures that are adjusted for the school context. Recent research on data use in 178 secondary schools in England in the United Kingdom shows overwhelmingly that teachers regularly use data (813 teachers completed surveys) (Kelly and Downey, 2011). An interesting finding was that schools with significantly high contextual value added scores reported both greater use of data and satisfaction with using data. The researchers suggest that this may indicate a developed culture of data use at both the student and school levels, given that contextual value added data show the degree of progress by every student – thus generating a need for more teachers to use data in order to make the task manageable. Similarly, the OECD Review in the Flemish Community of Belgium revealed appreciation by schools for the contextual value-added performance feedback from the national sample assessments. (It should be noted that these results are not published).

Box 6.20 The public reporting of school performance

Taking account of factors that impact student learning and are beyond the control of schools

In Sweden, the National Agency for Education has developed two public databases presenting comparable information on different schools (and local authorities). One includes basic statistical information as well as student test scores (SIRIS), the other presents statistical measures on how each school performs given its "expected value" as calculated in a regression model (SALSA). This notion of expected value is essentially an adjustment in a statistical regression model of the school's performance according to its particular student composition. The regression model takes into account: a weighted indicator on the parents' education; the proportion of boys; the number of students born abroad and the proportion of students born in Sweden with both parents born abroad. In this way, a given school with a large proportion of students with parents at lower educational level is assumed to perform less well compared to a school with a small proportion of such students. A comparison of the school's average student performance with the school's "expected value" calculated in the regression model (which corresponds to the average score for schools with the same student composition) represents a proxy of the value that school brings.

In Australia, the school reporting website (My School) uses a measure of socio-educational advantage to present "fair and meaningful" comparisons of school performance on the national assessments (NAPLAN). The Australian Curriculum, Assessment and Reporting Authority (ACARA) developed an Index of Community Socio-Educational Advantage via a statistical model that identified the combination of variables that were most closely associated with student achievement in the national assessments (NAPLAN). Background information for individual students (parental education and occupation) is aggregated to the school level and then combined with school community factors (proportion of indigenous students, remoteness of the school and a measure of educational disadvantage for students with a language background other than English). Using this statistical adjustment allows for a school's performance to be compared with statistically similar schools.

Showing the school's contribution to student learning progress

In Australia, students sit national assessments (NAPLAN) at four different year levels. The school reporting website (My School) uses a measure of "student gain" to present school performance on NAPLAN. This presents average results for students who sat the national assessments on two occasions at the same school and have results at two year levels. The percentage of students in the school that are included in this measure is also displayed. The student gain in a given school can be compared to: the Australian average student gain; average student gain in similar schools as measured in the contextual adjustment measure (see above); and student gain for all students across Australia who had a similar starting point.

In England in the United Kingdom, a performance table is reported on line for each school and includes information on the learning progress of students between different key stages of the national curriculum. The proportion of students in each school that makes "expected progress" is based on national minimal expected learning progress between two different key stages. In secondary schools, the national examinations at age 16 (General Certificate of Secondary Education) are used to measure performance at the end of Key Stage 4 and the best eight examination results are taken for each student to measure the "added value" since the end of Key Stage 2. The percentage of students in the school that are included in this measure is displayed and value added results are only shown for schools where this comprises at least 50% of students.

For further information, see: *http://salsa.artisan.se*; *www.education.gov.uk/schools/performance/*; *www.myschool.edu.au*.

Pointers for future policy development

This chapter has reviewed the approaches countries are taking to school evaluation in light of available research and evidence. The policy suggestions that follow are drawn from the experiences reported in the Country Background Reports, the analyses of external review teams in Country Reviews and the available research literature. It should be stressed that there is no single model or global best practice of school evaluation. The development of practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing policies for school evaluation for different countries.

In general, there is a need for further research into the impact of different policy approaches to school evaluation. The existing evidence base is dominated by research in a few systems with long-established policies in school evaluation. As more systems adopt and implement different school evaluation policies, there will be a need to collect evidence on how these impact student learning and educational experiences.

Governance

Clarify the role and purpose of school evaluation within the wider evaluation and assessment framework

School evaluation in any system must be seen in the context of its particular cultural traditions as well as the wider policy arena if its precise nature and purpose is to be understood. The OECD Review considers school evaluation as one of several key components in the overall evaluation and assessment framework for a school system. As such, its particular role and contribution should be aligned to the wider goals for the school system and considered in the overall balance of accountability and development functions within the evaluation system. The development of school evaluation will depend on a range of established practices in the school system such as the extent of school autonomy, the extent of market mechanisms and the culture of evaluation. As part of a general agenda, the fundamental purpose of school evaluation needs to be clearly and consistently understood across the school system. For instance, external school evaluation can be part of the strategy to bring about general improvement across all schools or, more narrowly, it can focus on "underperforming schools". The approach adopted depends on the underlying policy agenda and the evidence about the performance of the school system as a whole.

Meaningful school evaluation involves: an accurate assessment of the effectiveness of schools; an assessment of strengths and areas for development, followed by feedback, coaching, support and opportunities for development; an opportunity to celebrate, recognise and reward the work of schools and to identify best practice; and an opportunity to identify underperforming schools.

The purpose of different types of school evaluation should underlie all key decisions in designing a school evaluation framework. From the outset, the purpose of school evaluation is of critical importance in deciding: who should be responsible for undertaking the evaluation, which procedures should be used and how the results of the evaluation will

be used. Within a school evaluation framework, consideration should be paid to which elements of school evaluation best serve accountability purposes and which best serve development purposes. For example, there would be strong expectations for transparency in reporting the results of school evaluation which mainly serves accountability purposes and for such evaluations to be based on explicit evaluation criteria.

Further, given that school systems are dynamic and that student learning objectives may evolve, the school evaluation framework would need to adapt to meet demands for meaningful feedback against these changing objectives. This necessitates a firm research approach to regularly evaluate the impact of different evaluation approaches and their fitness for purpose.

Ensure the focus for school evaluation is the improvement of teaching, learning and student outcomes

School evaluation demands significant capacity at many levels of the school system. It is crucial to ensure that school evaluation contributes towards school improvement and is not perceived as an exercise in compliancy. The approach to school evaluation (both external school evaluation and school self-evaluation), the criteria and questions governing judgements and the methods employed should focus directly on the quality of teaching and learning and their relationship to student learning experiences and outcomes. This requires a culture of openness and reflection around what happens during the teaching and learning process, including classroom observation.

It is important that school evaluations do not focus simply on the relationship between policy, planning and outcomes. The most important contribution which school evaluation can and should make to understanding the performance of a school is its focus on teaching and learning. The quality of teaching is central to the quality of student's learning and the key variable which a school can influence. The central task of school evaluation, therefore, is to determine the quality of teaching across the staff as a whole. This can be a sensitive issue but sends the signal to students, teachers and parents that school evaluation is not a bureaucratic exercise which is largely the concern of school leadership but relates to the work of each and every member of staff.

Evaluate and adapt external school evaluation to reflect the maturity of the school evaluation culture

There is generally a need to have better research on the impact of different approaches to external school evaluation. A strong evidence base on the performance of the school system is essential in guiding decisions on how to allocate most effectively resources for external school evaluation. There may be demands to reduce the frequency of external school evaluations or the intensity of the evaluation visit in terms of length of time spent at the school. However, such decisions need to be based on a careful evaluation of the evidence of school performance and self-evaluation culture throughout the school system and need to ensure the continued legitimacy and respect by educators for the external school evaluation process.

External school evaluation approaches are changing in a number of systems with a move to a differentiated approach based on the assessment of risks to school quality in different schools. The idea behind this is to focus external school evaluation on the schools that need this most and sometimes in the specific pedagogical areas that need most attention. Systems adopting such a differentiated approach typically adopt a policy

ensuring the evaluation of all schools within an agreed time period (e.g. anything from five to ten years), but focus external school evaluations on schools where particular concerns have been identified against a desk-based assessment of risks (e.g. among others, parental complaints, high staff turnover, weak or weakening student outcomes). There could also be differentiation in terms of the focus of the school evaluation, i.e. emphasis on particular factors of concern in that school and not on the full set of factors identified in the national external school evaluation framework.

Moving to differentiated external school evaluation models requires a high level of intelligence about school characteristics and performance. Hence, it is recommended that systems move to this approach once the evaluation culture is consolidated, evaluation capacity in schools is satisfactory and data gathering and analysis within the school evaluation framework is established. Of particular importance in moving to a differentiated approach is to ensure that schools that are not identified for external school evaluation (schools judged to be of low risk and good quality) do not become complacent. Policies, therefore, need to establish requirements for low risk schools to provide evidence about progress on a broad front.

Raise the profile of school self-evaluation

School self-evaluation is of key importance to school improvement and quality assurance and needs to be consolidated in school systems. An option to strengthen self-evaluation is to establish requirements for schools that promote strategic planning, for example, the drawing up of a 4-to-5 year strategic plan and regular updates of school progress on this plan, or the development of annual school reports about their achievements, challenges and strategies for improvement. The process of meeting specified strategic planning requirements would be a stimulus for many schools to further their self-evaluation practices and would hold strong potential for school improvement, if: the reporting and planning pays sufficient attention to key processes of teaching and learning and a broad range of outcomes; the process of reporting and planning adequately engages the school community; and the school community takes keen interest in school progress towards its strategic goals. For example, when establishing an annual strategic plan, schools would determine priorities for action over the year, set their own targets in line with local needs and priorities and decide on the assessment methods to monitor progress. This approach allows schools to take responsibility for their own improvement strategies. There is a note of caution on setting specific requirements for schools to publish self-evaluation results, as this may hinder their use for school development. The essential aim is that the school community is engaged in the process, owns the process and makes use of the results to continually strive for improvement in teaching and learning in the school.

There is also a role for external school evaluation to promote the reinforcement of school self-evaluation practices. External school evaluations have the potential to build capacity in schools for school-based self-evaluation and will increase evaluation literacy in schools. Schools may be motivated to engage in self-evaluations if faced with an external school evaluation requirement, even when school self-evaluation is not suggested as an alternative to external school evaluation but only as a prior condition and counterpart. Further, external school evaluations may promote a more formalised and extended process of self-evaluation in schools. Schools may become more willing to use methods of evaluation that had not necessarily been used previously. External school evaluations can bring greater depth and breadth to self-evaluations in schools when they for example provide the school with relevant benchmark information, comparative data

from other schools or new and challenging ideas that might help the school to expand its evaluation, interpret its own data and assess its quality.

Align external school evaluation with school self-evaluation

A combination of school self-evaluation and external school evaluation can maximise the benefits of both while counteracting the limitations arising from an over reliance on the use of only one. A reliance on external school evaluation alone can promote a culture of compliance or "gaming" within which schools seek to satisfy the demands of external school evaluation but fail to take ownership of or accept responsibility for improvement. Self-evaluation is integral to continuous improvement which is not solely reliant on the impact of external school evaluation. However, self-evaluation can also be subject to self delusion where assumptions are not challenged and power relationships in the school community have an undue influence on what is evaluated and the nature of the judgements themselves.

As a result, good alignment is needed between policy and practice in both external school evaluation and school self-evaluation. In particular, there is a need to ensure that the criteria used in both spheres are sufficiently similar as to create a common language about priorities and about the key factors which influence high-quality teaching and learning. Lack of clarity about what matters is likely to relegate self-evaluation to something which serves external school evaluation rather than creating a platform for an exchange based on reliable and comparable evidence. Other strategies to ensure the alignment between external school evaluation and school self-evaluation include giving a strong focus on how the school is going about its own self-evaluation and using the results to improve learning; and collaborating with schools to validate their self-evaluations and the steps they are taking to bring about improvement.

The priority is to ensure that school self-evaluation and external school evaluation are complementary and mutually reinforcing processes. The basic premise is that schools are best placed to analyse their own contexts and that external school evaluation can provide an external perspective to validate or challenge the schools' own findings.

Procedures

Develop nationally agreed criteria for school quality to guide school evaluation

The coherence of school evaluation is considerably enhanced when based on a nationally agreed model of school effectiveness. This national model should draw on both international and national research that has identified the factors generally associated with the quality of teaching and learning. This would provide clear criteria for effective schools and provide a robust, research-based foundation for both school self-evaluation and external school evaluation. Such criteria would form the basis of any external school evaluation framework, e.g. a national inspection framework. Further, schools would use these criteria and benchmarks to consider the evidence needed to rate their own effectiveness. "How good is our school?" is a central question not only for students and parents but for those who lead and work in schools. Similarly, "How good are our schools?" is the question for educational authorities.

An agreed framework of school quality indicators should be established, which could then be made widely available to schools and school organisers to use in their own evaluative processes. This will increase the alignment between school self-evaluation and external school evaluation, which has the advantage of keeping schools focused on core

quality criteria in a more systematic fashion and not just in relation to cycles of external school evaluation. School quality indicators will address contextual, input and process factors, but should put focus on a broad range of student outcomes. For example, quality indicators for student outcomes and their rate of progress could include the extent to which every student in a school: is making better than expected progress given their earlier attainment; is pleased with the education at the school; feels safe and happy at school; gains the knowledge, skills, understanding and attitudes necessary for lifelong fulfilment, etc.

Finally, there should be a periodic evaluation of the school quality indicator framework and criteria to ensure it reflects updated evidence from research and stakeholders on the factors associated with the quality of teaching and learning.

Develop appropriate resources for school self-evaluation

There is a role for systems to offer schools self-evaluation resources and tools. Access to consistent, comparable, reliable and broad-based self-evaluation tools and examples of effective use of these in school policy making would give school principals a better picture of what school self-evaluation looks like when it is working well. This is also a way to promote the collection of more qualitative evidence by schools in their self-evaluation. Stakeholder surveys are already an established feature of school evaluation in a number of systems and are increasingly a requirement of school reporting. While particular instruments are not always mandatory, the principle of gathering evidence about perceptions and levels of satisfaction is now an expectation in some systems. Further, there is room to centrally promote examples of where schools are working effectively with self-evaluation tools. The efficient feedback of key centrally collected information to the school level also plays an important role (see below).

Ensure a strong evidence base for external school evaluation and appropriate analysis tools

Credible external school evaluation should be based on reliable and relevant evidence rather than opinion. Acceptance of external school evaluation results can be secured through systematic gathering, analysis of and reference to relevant evidence. An effective way to pull together key information is to compile a school profile, comprising key school quality indicators. Providing this to an external school evaluation team prior to the evaluation aids efficiency by allowing the team to focus its attention on key issues. Further, the school profile can help to benchmark and contextualise the evaluators' judgements. Such a profile is particularly helpful when based on robust and comparable evidence on school outcomes. This is a critical element in a system of external school evaluation that relies on a form of risk assessment to determine the cycle and focus of external school evaluations.

Similarly, evidence should be collected during the course of external school evaluation, including the identification and analysis of documentation, the collection of feedback on school quality via stakeholder surveys, and interviews with a representative sample of stakeholders.

A key part of external school evaluation is the observation of classroom teaching and learning. This necessitates high levels of skill in the techniques of observation and appropriate training. But the objectivity of observations can also be enhanced by the development of observation indicators and specific training on the use of these indicators. Similar instruments can be developed to support the decision-making process of external

school evaluators during school visits. These can identify key criteria and clarify the rules on forming judgements, by providing examples of how different observed phenomena would be rated.

Ensure transparency in external school evaluation procedures

The principle of transparency is increasingly perceived as an integral part of effective external school evaluation. Such transparency in the methodology, process and results of external school evaluation is perceived as being fairer to those evaluated and a way to promote the integrity, rigour and impact of external school evaluation. The approach, procedures and instruments used in external school evaluation are now routinely available on the Internet for public consultation and external school evaluation reports themselves are published either in paper form or digitally. Typically, the criteria for evaluation and the evaluation instruments are publicly available and the evaluation team actively encourages the school to examine this documentation in advance.

Another important aspect of transparency is to include processes allowing schools to comment on their experience with external school evaluation. External evaluators can systematically seek feedback from schools on their experience with the external school evaluation process. Importantly, schools should also be given the possibility to comment on the evaluator's report. For example, schools should be able to correct factual errors and to challenge findings in the evaluator's report. This could even include allowing schools the possibility to include an official statement on the evaluation findings in the published report.

To ensure that external school evaluation results are taken seriously by schools, there should be clearly defined procedures on how evaluation results will be followed up by schools and the external school evaluation bodies, including where necessary timelines for improvement and consequences for inadequate improvement (see below).

Developing school evaluation capacity – a priority for school improvement

Ensure the credibility of external evaluators and enhance their objectivity and coherence

The selection and recruitment of external evaluators is of key importance in building capacity within the external school evaluation body. The criteria used to select evaluators should be demanding to ensure that those recruited have the skills and attributes necessary for a credible approach to external school evaluation. Externality implies sufficient distance from responsibility for the school's performance to avoid conflicts of interest and perceived bias. The range of individuals who are part of external school evaluation teams should also be broad. The use of highly credible school principals and leading practitioners in external school evaluation would both heighten the credibility of the evaluation teams and build capacity in the school system as a whole.

In addition to offering specific training for external evaluators, external school evaluation can also be organised in ways that enhance the coherence of evaluators' judgements. Examples include the use of the same evaluation teams in a common group of schools or the organisation of regular meetings of external evaluators within the external school evaluation body.

External school evaluation bodies should implement internal mechanisms to regularly evaluate the coherence and quality of external school evaluation procedures. Importantly,

there should be mechanisms to seek feedback from key stakeholders on their experience with the external school evaluation. Such information can form the basis of identification and analysis of ways to improve the external school evaluations. Further, this information is particularly useful when implementing a new approach to external school evaluation.

Ensure sufficient capacity and retraining as necessary to fit the approach to external school evaluation

Governance decisions on the approach to external school evaluation will directly impact the required capacity for external school evaluation. This may involve the introduction or reintroduction of a system of external school evaluation, which would require establishing an external school evaluation body. Such decisions have significant resource implications. The adequate resourcing and provision of training to a new external school evaluation body will play a crucial role in building its reputation among schools. In another scenario, there may be a need to reduce the capacity of the external evaluation body and this would have implication for the frequency and/or intensity of external school evaluations. With the strengthened role for school self-evaluation within the school evaluation framework, external school evaluators need to update their skills to be able to validate school self-evaluation and even to work collaboratively with schools on their school self-evaluations.

Strengthen school principals' capacity to stimulate an effective school self-evaluation culture

There needs to be an explicit recognition that the process of self-evaluation is hugely dependent on school leadership's capacity to stimulate engagement, to mobilise resources and to ensure appropriate training and support. The drawing up of national and/or professional school principal and deputy principal competency profiles should clarify the importance of the school self-evaluation process, including classroom observation in the school principal's role. Attention should also be paid to ensuring adequate training opportunities are available to school principals in these key areas. School self-evaluation can be promoted by training school principals in school effectiveness and its evaluation, including the techniques of observing and assessing teaching and learning and giving developmental feedback. It is essential to ensure that school principals and other members of the school with evaluation responsibilities have the necessary skills in class observation, interviewing, data gathering, analysis and interpretation of results which both ensure validity and reliability in the evaluation process and which allow the results of evaluation to be understood. Consideration can also be given to the resourcing of structures to strengthen school principals' capacity to implement effective self-evaluation processes, for example, by creating new evaluation roles within the school for different staff.

Promote the engagement of all school staff and students in school self-evaluation

School self-evaluation activities should not remain an exercise for the school leadership team, but should engage the school staff and students. There is considerable recognition of the importance of fully engaging all members of the school community in the self-evaluation process. However, there is also evidence that this requires high levels of trust and strong commitment from the school community.

Students have important feedback to give to their schools. Evidence from several systems has highlighted that involving students in decisions about their schooling is an important factor in school improvement. There are several approaches to engaging

students' feedback, from establishing student councils, through the use of student surveys in schools, to involving students in the feedback to teachers on their teaching.

There is also a need to focus on helping school staff interpret and translate evaluative information into action. School evaluation will not lead to improvements unless the information gathered is interpreted and translated into strategies for school development. Substantial investment needs to be directed at strategies to ensure that professionals are able use the feedback they receive effectively (see also above).

Promote peer learning among schools

In all systems, there is much potential for schools to collaborate and learn from each other in the process of evaluating and improving processes and outcomes. This is a particularly useful strategy in systems where there is a high degree of school autonomy, as it can prevent schools from forming an introspective and defensive culture. Leadership standards in a number of systems highlight the importance of networking and partnerships between schools. Providing funding for groups of schools to work collaboratively would provide an incentive and stimulate collegial networking, peer exchange, sharing and critiquing of practice, fostering a sense of common direction. Critical friendship does not just happen by chance. It needs development, including the development of observation and evaluation skills, and skills of professional dialogue. It also requires the development of trust. A starting point could be with school leadership teams working together to identify common challenges, devising common strategies and approaches to peer school evaluation. The process would benefit from the appointment of an external facilitator or critical friend chosen and agreed by the school principals themselves. Within systems, there are schools with more developed self-evaluation processes and there could be great benefits in finding ways to involve their staff in supporting and training colleagues in other schools.

Reporting and use of results

Optimise the feedback of nationally collected data to schools for self-evaluation and development planning

The administration and collection of results from national student assessment programmes represents an important investment. It is, therefore, critical that systems are in place to optimise the reporting and feedback of results to schools. There are different levels of decisions here. First, any concerns on confidentiality of data. Second, ways to feedback results to different levels to optimise their use for improvement. For example, school principals will benefit from an overview of results for the school and also from comparative performance information against other schools, regions or national averages. Teachers will benefit from the feedback of information at the class level and individual student level, as useful diagnostic evidence. Third, the timeliness of the feedback of results is a key consideration. The faster the feedback of student results to teachers, the more relevant they are for adapting instructional practices in particular classes or with particular students.

Technology offers opportunities to enhance both the nature and timeliness of feedback. The speed of feedback has major implications in the choice of assessment medium: results from computer-based tests can be more readily compiled, scored and reported back to teachers and schools. Reporting back results via electronic portals can capitalise on the ability to set confidential access for different users to different reports

and to provide users with analytical software to select and compare performance of different tailor-made groupings.

School self-evaluation efforts will also benefit from the ready access to centrally held information, for example as reported against external school evaluation frameworks.

Promote the wider use of the results of external school evaluation

The publication of all external school evaluation reports is associated with many benefits. The school community can use this information to feed into school development planning and there is emerging evidence that a keen interest from the school community in the results of external school evaluation is associated with school improvement actions. The publication of reports has also promoted a more coherent format of external school evaluation reports. This makes the information more helpful for schools to compare their external evaluation with other schools and can provide useful input to school self-evaluation activities.

External school evaluation reports should not be too technical and should be readable to a non-specialist audience. Improving the communication of external school evaluation results to a wider audience offers the opportunity to examine the terminology used within the external school evaluation framework. There may be room to make the framework more readily accessible to teachers and students at the same time as aiding the communication of results to the public. This could also enhance the alignment of external school evaluation and school self-evaluation.

It is important to develop a communication strategy that capitalises on the wider dissemination of school evaluation results. This could include different elements ranging from specific summaries for parents within the external school evaluation reports, through the publication of results for a group of schools within a particular area or educational group, to tailor-made websites enabling parents to consult reports for a given school and to compare particular aspects of that school with other schools in the local area or nationally. There is also a role for schools to be proactive in promoting external school evaluation results to staff and parents.

Ensure the systematic follow-up of external school evaluations

To heighten the impact of school external school evaluation on school improvement there needs to be systematic follow-up by the external evaluators and/or appropriate authorities or support agencies. Such follow-up should include both a monitoring and support function. Of course the starting point is to ensure that external school evaluation results in a good amount of feedback to schools, including a useful and practical level of detail on required improvements. In turn, this needs to be accompanied by the appropriate investment in strategies to ensure that schools effectively use the feedback they receive. The extent of follow-through activities by external evaluators and/or appropriate authorities could be made dependent on the extent of improvement needed by a school and its capacity to improve. In such a case, schools would benefit from a clear set of follow-up procedures, including for example the amount of time schools have to demonstrate their implementation of improvement plans and possibly requirements for schools to use external support in this process and clear criteria for when external support would be judged necessary. Further, there should be clear procedures in place for the further follow-up of schools that are judged not to have made adequate improvement upon a second external school evaluation.

Report a broad set of school performance measures with adequate contextual information

In systems where comparative national assessment data are published for individual schools, there is a strong case to provide complementary evaluative information such as external school evaluation reports which broaden the base of evidence and provide more explanation of the factors which have influenced school performance. Also, policy makers might lessen the potential undesired effects of the publication of test data by ensuring that quantitative data are always accompanied by a description of the context in which different schools operate; providing interpretation of data gathered in student assessments, school self-evaluations, and external school evaluations; describing how schools are meeting local goals for education, noting progress made in meeting challenges, describing new programmes under development, and so on.

The development of measures that adjust for students' prior attainment is widely supported (value added performance measures). However, these are not without considerable methodological challenges. There needs to be a balance between an attempt to present a fair comparison and ensuring that measures are not obscure and can be easily understood and interpreted by users. Such challenges hold true for measures to adjust for the school context, also. But these appear to be highly appreciated by schools with more challenging intake and can be very helpful in school self-evaluation as they allow schools to benchmark their results with other similar schools.

Notes

1. Although in this case the working group drew heavily on the "proportionate" inspection approach already used by the Scottish Inspectorate. In turn, recommendations from the Reducing Burdens Action Group impacted the Scottish Inspectorate's approach to its supervision of education providers to ensure it is: "focused on outcomes; proportionate to need; owned by those carrying out the self-evaluation; flexible, with the scope to recognise differences in service levels and types; built on existing good practice and relevant existing standards; rigorous and transparent; designed to secure continuous improvement".

2. The path analysis model has a good fit to the theoretical relations. The Comparative Fit Index (CFI) is 0.965 (this can range from 0 to 1 with a larger value indicating a better model fit and an acceptable fit being a value of 0.90 or greater [Hu and Bentler, 1999]); the Root Mean Square Error of Approximation (RMSEA) is 0.027 (this is related to residual in the model and can range from 0 to 1, with a lower value being a better fit and an acceptable fit being a value of 0.06 or less [idem]).

References

Allerup, P., V. Kovac, G. Kvåle, G. Langfeldt and P. Skov (2009), "Evaluering av det nasjonale kvalitetsvurderingssystemet for grunnopplæringen", Agderforskning, Kristiansand, Norway.

Alvik, T. (1996), *School Self-Evaluation: A Whole School Approach*, CIDREE, Dundee.

Ball, S. (2003), "The teacher's soul and the terrors of performativity", *Journal of Education Policy*, Vol. 18, No. 2, pp. 215-228, Routledge, London.

Barber, M. (2004), "The virtue of accountability: System redesign, inspection, and incentives in the era of informed professionalism", *Journal of Education,* Vol. 185, No. 1, *www.bu.edu/journalofeducation/files/2011/06/BUJOE-185.1Barber.pdf.*

Barber, M. (1996), *The Learning Game: Arguments for an Education Revolution*, Victor Gollanez, London.

Bekkers, V., D. de Kool and G. Straten (2012), "Beleidsgericht Onderzoek Primair Onderwijs", *Ouderbetrokkenheid bij Schoolbeleid*, pp. 108.

Blok, H., P. Sleegers and S. Karsten (2008), "Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model", *Journal of Education Policy*, Vol. 23(4), pp. 379-395.

Blondin, C. and B. Giot (2011), *Cadres d'évaluation en vue d'améliorer les resultants scolaires – Etude thématique de l'OCDE: Rapport de base pour la Communauté française de Belgique*, Ministère de la Fédération Wallonie-Bruxelles, *Brussels, www.oecd.org/edu/evaluationpolicy.*

Bolam, R., A. McMahon, L. Stoll, S. Thomas and M. Wallace, with A. Greenwood, K. Hawkey, M. Ingram, A. Atkinson and M. Smith (2005), *Creating and Sustaining Effective Professional Learning Communities*, Research Report 637, London DfES and University of Bristol.

Boyle A., K. Carlson Le Floch, S. Bowles Therriault and B. Holzman (2009), *State Support for School Improvement: School-level Perceptions of Quality – Evaluating the Quality of State Support for School Improvement*, American Institutes for Research, Washington.

Brejc, M., M. Sardoč and D. Zupanc (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Slovenia*, National School for Leadership Slovenia, *www.oecd.org/edu/evaluationpolicy.*

Chairmanship of the School Council / Skolerådet (2009), *Beretning om Evaluering og Kvalitetsudvikling af Folkeskolen 2009*, Formandskabet for Rådet for Evaluering og Kvalitetsudvikling af Folkeskolen, Copenhagen.

Chapman, C. (2005), *Improving Schools through External Intervention*, Continuum, London.

Chapman, C. and M. Hadfield (2010), "School-based networking for educational change", in A. Hargreaves, A. Lieberman, M. Fullan and D. Hopkins (eds.), *Second International Handbook of Educational Change*, Springer, Dordrecht.

Chvál, M., L. Procházková and K. Černý (2010), "Postoje k autoevaluaci ředitelů základních a středních škol v ČR" (Positions on self-evaluation of head teachers at basic and secondary schools in the CR), *Orbis Scholae*, Vol. 4, Issue 1, pp. 91-110.

Chvál, M. and K. Starý (2008), "Autoevaluace škol v reflexi poskytované odborné podpory" (Self-evaluation of schools from the perspective of expert assistance), *Orbis Scholae*, Vol. 2, Issue 3, pp. 53-73.

Commissie Monard (2009), "Kwaliteit en kansen voor elke leerling – Een visie op de vernieuwing van het secundair onderwijs, Voorstel van de Commissie Monard" (Quality and opportunities for every student – An opinion about the revision of secondary education, proposal of the Committee-Monard), *www.ond.vlaanderen.be/nieuws/2009/bijlagen/0424-visienota-SO.pdf.*

Corbett, H.D. and B.L. Wilson (1991), "Two state minimum competency testing programs and their effects on curriculum and instruction", in R.E. Stake (ed.), *Advances in Program Evaluation: Vol. I. Effects of Mandated Assessment on Teaching*, JAI Press, Greenwich, Connecticut, pp. 7-40.

Creemers, B.P.M. (2007), "Educational effectiveness and improvement: The development of the field in mainland Europe", in T. Townsend (ed.), *International Handbook of School Effectiveness and Improvement*, Springer, Dordrecht.

Creemers, B.P.M., L. Stoll, G. Reezigt and the ESI Team (2007), "Effective school improvement – ingredients for success: The results of an international comparative study of best practice case studies", in T. Townsend (ed.), *International Handbook of School Effectiveness and Improvement*, Springer, Dordrecht.

Croxford, L., S. Grek and F.J. Shaik (2009), "Quality assurance and evaluation (QAE) in Scotland: Promoting self-evaluation within and beyond the country", *Journal of Education Policy*, Vol. 24, No. 2, March 2009, pp. 179-193.

Danish Ministry of Education and Rambøll (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Denmark*, Agency for the Evaluation and Quality of Primary and Lower Secondary Education, *www.oecd.org/edu/evaluationpolicy.*

Dedering, K. and S. Müller (2010), "School improvement through inspections?" First empirical insights from Germany, *Journal of Educational Change*, Vol. 12, Issue 3, Springer, pp. 301-322.

Department of Education and Science (2005), *Customer Survey* – A report of a Customer Survey by MORI Ireland on behalf of the Inspectorate of the Department of Education and Science, Ireland.

Department of Education and Science (1999), *Whole School Evaluation Report on the 1998/1999 Pilot Project*, Ireland, *www.education.ie/servlet/blobservlet/des_wse_report_98_99_pilot_prj.pdf.*

Department of Public Expenditure and Reform (2012), *Third Report of the Organisational Review Programme (ORP)*, Dublin, Ireland, *www.taoiseach.gov.ie/eng/Publications/Publications_2012/The_3rd_Report_of_the_ORP.pdf*.

Döbert H., E. Klieme and W. Sroka (eds), (2004), *Conditions of School Performance in Seven Countries: A Quest for Understanding the International Variation of PISA Results*, Waxmann Verlag GmbH.

Dos Santos, S. and R. Rakocevic (2012), *Les cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude de l'OCDE : Rapport national de base pour la France*, Direction de l'Évaluation, de la Prospective et de la Performance (DEPP), Ministère de l"Éducation nationale, *www.oecd.org/edu/evaluationpolicy*.

Earl, L., N. Watson, B. Levin, K. Leithwood, M.Fullan and N. Torrance (2003), *Watching and Learning 3: Final Report of the External Evaluation of England's Literacy and Numeracy Strategies: Executive Summary*, Department for Education and Skills, Nottingham.

Education Scotland (2011), *Arrangements for Inspecting Schools in Scotland* – August 2011, Education Scotland, *www.educationscotland.gov.uk/Images/SchoolInspectionFramework2011_tcm4-684005.pdf*.

Ehren, M.C.M., H. Altrichter, G. McNamara and J. O'Hara (2013), "School inspections and school improvement: Testing assumptions on causal mechanisms", *Oxford Review of Education* (submitted and under review).

Ehren, M.C.M., H. Altrichter, G. McNamara and J. O'Hara (2012), "Impact of school inspections on improvement of schools – describing assumptions on causal mechanisms in six European countries", *Educational Assessment, Evaluation and Accountability*, January 2013, Springer.

Ehren, M. and M. Hendriks (2010), *Combinations of Internal Evaluations and External Accountability in Europe: Ideal Types and Existing Models*.

Ehren, M.C.M. and A.J. Visscher (2008), "The relationship between school inspections, school characteristics and school improvement", *British Journal of Educational Studies*, Vol. 56, No. 2, June 2008, Blackwell Publishing Ltd, Oxford, pp. 205–227.

Eivers, E., S. Close, G. Shiel, D. Millar, A. Clerkin, L. Gilleece and J. Kiniry (2010), *The 2009 National Assessment of Mathematics and English Reading in Primary Schools*, *www.erc.ie/documents/na2009_report.pdf*.

Ekonomistyrningsverket (2006), "Skolverkets utbildningsinspektion – ger den några effekter? Effektutvärdering genomförd på uppdrag av Skolverket", Ekonomistyrningsverket (Swedish National Financial Management Authority), Stockholm.

Englert, K., D. Fries, M. Martin-Glenn and B. Douglas (2007), "Accountability systems: A comparative analysis of superintendent, principal and teacher perceptions", *International Journal of Education Policy and Leadership*, Vol. 2 (4).

ERO (Education Review Office) (2011), *Framework for School Reviews 2011*, Education Review Office, New Zealand, *www.ero.govt.nz/Review-Process/Frameworks-and-Evaluation-Indicators-for-ERO-Reviews*.

ERO (2010), *Preparing to Give Effect to the New Zealand Curriculum*, Education Review Office, Wellington, *www.ero.govt.nz/National-Reports/Preparing-to-Give-Effect-to-the-New-Zealand-Curriculum-June-2010.*

ERO (2007), *The Collection and Use of Assessment Information in Schools*, Education Review Office, Wellington.

European Commission (2011), *The Making of: Leadership in Education*, A European Qualification Network for Effective School Leadership, Niedersächsiches Landesinstitüt für schulische Qualitätsentwicklung (NLQ), Hildesheim.

Eurydice (2012), *Key Data on Education in Europe 2012*, Education, Audiovisual and Culture Executive Agency, Eurydice, Brussels, *http://eacea.ec.europa.eu/education/eurydice*.

Eurydice (2004), *Evaluation of Schools Providing Compulsory Education*, Eurydice, Brussels.

EVA (Danish Evaluation Institute) (2010), *Strategies for Development of Leadership in Daycare Centres and Schools*, Danmarks Evalueringsinstitut, Copenhagen.

EVA (2009), *Kommunernes arbejde med kvalitetsrapporter*, Danmarks Evalueringsinstitut, Copenhagen.

EVA (2007), *Kvalitetsarbejde: erfaringer fra 20 skoler*, Danmarks Evalueringsinstitut, Copenhagen.

Faubert, V. (2009), "School evaluation: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 42, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Figlio, D. and S. Loeb (2011), "School accountability", in E. Hanushek, S. Machin and L. Woessman (eds.), *Handbooks in Economics*, Vol. 3, North-Holland, The Netherlands, pp. 383-421.

Flemish Ministry of Education and Training (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, Flemish Ministry of Education and Training, Brussels, *www.oecd.org/edu/evaluationpolicy*.

Fournier, G. and D. Mildon (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Canada*, Council of Ministers of Education, Canada (CMEC).

Grillitsch, M. (2010), "Evaluation der Implementation von Bildungsstandards. Evaluation der Ergebnisrückmeldung. Erste Ergebnisse der Befragung von Lehrer/innen und Schulleiter/innen der Baselineschulen (März 2010)", unpublished manuscript, BIFIE, Graz.

Haider, G. (2006), "Schulqualität und Qualitätsentwicklung", in G. Haider and C. Schreiner (ed.), *Die Pisa Studie. Österreichs Schulsystem im internationalen Wettbewerb*, Böhlau, Wien, pp. 353-366.

Hamilton, L. and D. Koretz (2002), "Tests and their use in test-based accountability systems", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education*, RAND Publishing, Santa Monica, California.

Harris, A. and J. Chrispeels (2006), *Improving Schools and Educational Systems*, Routledge, Oxford.

Her Majesty's Inspectorate of Education (HMIE) (2007), *How Good is our School? The Journey to Excellence: Part 3*, HMIE, Livingston, *www.educationscotland.gov.uk/Images/HowgoodisourschoolJtEpart3_tcm4-684258.pdf*.

Hill, R. and P. Matthews (2010), *Schools Leading Schools II: The Growing Impact of National Leaders of Education*, National College for Leadership of Schools and Children's Services, Nottingham, England.

Hofman, R.H., N.J. Dukstra and W.H.A. Hofman (2009), "School self-evaluation and student achievement", *School Effectiveness and School Improvement*, Vol. 20, No. 1, pp. 47-68.

Hofman, R.H., N.J. Dukstra and W.H.A. Hofman (2005), "School self-evaluation instruments: An assessment framework", *Leadership in Education*, July-September 2005, Vol. 8, No. 3, pp. 253-272.

Hooge, E., T. Burns and H. Wilkoszewski (2012), *Looking Beyond the Numbers: Stakeholders and Multiple School Accountability*.

Hopkins, D. (1995), "Towards effective school improvement", *School Effectiveness and School Improvement*, Vol. 6, Issue 3, pp. 265-274.

Hoyle, R. and J. Robinson (2003), "League tables and school effectiveness: A mathematical model", *Proceedings of the Royal Society of Biological Science*, Vol. 270, No. 1511, pp. 113-119.

Hu, L. and P.M. Bentler (1999), "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives", *Structural Equation Modeling*, Vol. 6(1), pp. 1-55.

Hutchinson, C. and M. Young (2011), "Assessment for learning in the accountability era: Empirical evidence from Scotland", *Studies in Educational Evaluation*, Vol. 37, Elsevier, pp. 62-70.

Inspectie van het Onderwijs (2012), *Toezichtkader PO/VO 2012 – Versie 1 juli 2012*, Inspectie van het Onderwijs, Utrecht, *www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2012/06/11/bijlage-bij-de-brief-van-minister-bijsterveldt-vliegenthart-ocw-aan-de-tweede-kamer-over-het-toezichtkader-po-vo-2012-van-de-inspectie-van-het-onderwijs.html*.

Inspectie van het Onderwijs (2010a), *Tevredenheidsonderzoek Periode juni–oktober, 2010*, Utrecht.

Inspectie van het Onderwijs (2010b), *Risico gericht toezicht beoordeeld*, Utrecht.

Inspectie van het Onderwijs (2009), *International Comparative Analysis of Learning and Teaching in Math Lessons in Several European Countries*, Utrecht, *www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2009/ICALT.pdf*.

Inspection Council Bureau (2009), *Inspection Reform Programme – More Effect, Less Burden*, Inspection Council Bureau, The Hague.

Irish Department for Education and Skills (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Ireland*, Department for Education and Training, Dublin, *www.oecd.org/edu/evaluationpolicy*.

Jackson, D. and J. Temperley (2007), "From professional learning community to networked learning community", in L. Stoll and K.S. Louis (eds.), *Professional Learning Communities: Divergence, Depth and Dilemmas*, Open University Press/McGraw Hill, Maidenhead.

Janssens, F.J.G. (2012), "The evaluation of the impact of the publication of school performance indicators in the Netherlands", *Journal of Education Policies*, Vol. 6(1).

Janssens, F. and G. van Amelsvoort (2008), "School self-evaluations and school inspections in Europe: An exploratory study", *Studies in Educational Evaluation*, No. 34, Elsevier, pp. 15-23.

Jung, T., J. Namgoong and J. Kim (2008), *Synthetic Report of School Assessment*, Korean Educational Development Institute, Seoul.

Katsuno, M. and T. Takei (2008), "School evaluation at Japanese schools: Policy intentions and practical appropriation", *London Review of Education*, Vol. 6, No. 2, July 2008, pp. 171-181.

Kelly, A. and C. Downey (2011), "Professional attitudes to the use of pupil performance data in English secondary schools", *School Effectiveness and School Improvement*, Vol. 22(4), pp. 415–437.

Kertesi, G. (2008), "A közoktatási intézmények teljesítményének mérése-értékelése, az iskolák elszámoltathatósága" (Assessment and Evaluation of the Performance of Public Education Institutions, Accountability of Schools), in K. Fazekas, J. Köllő and J. Varga (eds), *Zöld könyv a magyar közoktatás megújításáért* (Green Book for the Renewal of Hungarian Public Education), Ecostat, Budapest. pp. 167-189.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and Jungha Park (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

Kim, S., T. Jung and J. Kim (2009), *Analytical Study on School Evaluation Results and Improvement Measures*, Korean Educational Development Institute, Seoul.

Kogan, M. and M. Maden (1999), "An evaluation of evaluators: The OFSTED system of school inspection", in C. Cullingford (ed.), *An Inspector Calls; Ofsted and its Effect on School Standards*, Kogan Page Limited, London, pp. 9–32.

Koretz, D. (2010), *Implications of Current Policy for Educational Measurement*, Education Testing Service (ETS).

Kyriakides, L. and R.J. Campbell (2004), "School self-evaluation and school improvement: A critique of values and procedures", *Studies in Educational Evaluation*, Vol. 30, pp. 23-36.

Langfeldt, G., E. Elstad and S. Hopmann (eds.) 2008, "Ansvarlighet i skolen. Politiske spørsmål og pedagogiske svar. Resultater fra forskningsprosjektet" (Achieving School Accountability in Practice), Cappelen akademisk forlag, Norway.

Leithwood, K. and R. Aitken (1995), *Making Schools Smarter: A System for Monitoring School and District Progress*, Corwin, California.

Leithwood, K., C. Day, P. Sammons, A. Harris and D. Hopkins (2006), *Seven Strong Claims about Effective School Leadership*, The National College for Leadership of Schools and Children's Services, Nottingham, England.

Livingston, K. and J. McCall (2005), "Evaluation: Judgmental or developmental?" *The European Journal of Teacher Education*, Vol. 28, No. 2.

MacBeath, J. (2008), "Leading learning in the self-evaluating school", *School Leadership and Management*, Vol. 28, No. 4, September 2008, pp. 385-399.

MacBeath, J., B. Boyd, J. Rand and S. Bell (1996), *Schools Speak for Themselves*, reported commissioned by the National Union Teachers in the UK.

MacBeath, J. with M. Schratz, D. Meuert and L. Jakobsen (2000), *Self-Evaluation in European Schools*, RoutledgeFalmer, London.

Madaus, G. (1988), "The influence of testing on the curriculum", in L. Tanner (ed.), *Critical Issues in Curriculum: 87th Yearbook of the NSSE*, Part 1, University of Chicago Press, Chicago, pp. 83-121.

Maes, B., E. Ver Eecke and M. Zaman (2002), *School Self-Evaluation and Inspection: A Review of the Literature*, SICI, Belfast.

Matthews, P. and P. Sammons (2004), *Improvement through Inspection: An Evaluation of the Impact of Ofsted's Work*, Crown Copyright.

McDonnell, L.M. and C. Choisser (1997), *Testing and Teaching: Local Implementation of New State Assessments*, CSE Tech. Rep. 442, Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

Ministère de l'Éducation nationale (2009), *Mission des corps d'inspection: inspecteurs d'académie-inspecteurs pédagogiques régionaux et inspecteurs de l'Éducation nationale affectés dans les académies*, Bulletin officiel no. 22 du 28 mai 2009, *www.education.gouv.fr/cid28292/mend0910498c.html.*

National Audit Office (2009), *Partnering for School Improvement*, NAO, London, *www.nao.org.uk.*

Nees, P. (2007), *Schools and their ERO Recommendations: A Study of Six Wellington Area Schools*, *www.educationalleaders.govt.nz/content/download/744/6082/file/nees-sabbatical-06.pdf*.

New Zealand Ministry of Education (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for New Zealand*, Ministry of Education, New Zealand, *www.oecd.org/edu/evaluationpolicy*.

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, Norwegian Directorate for Education and Training, Oslo, *www.oecd.org/dataoecd/7/10/47088605.pdf*.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011b), *OECD Review of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

OECD (2012), *Teaching Practices and Pedagogical Innovation: Evidence from TALIS*, OECD Publishing, Paris.

OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, Paris.

OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

OECD (2008), *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*, OECD Publishing, Paris.

Oterkiil, C. and S. K. Ertesvag (2012), "Schools' readiness and capacity to improve matter", *Education Inquiry*, Vol. 3, No. 1, March 2012, pp.71-92.

Ouston, J., B. Fidler and P. Earley (1997), "What do schools do after OFSTED school inspections – or before?", *School Leadership and Management*, Vol. 17, No. 1, pp. 95–104.

Padilla, C., K. Woodworth and K. Laguarda (2006), "Evaluation of title I accountability systems: School-improvement efforts and assistance to identified schools", paper presented at the 2006 annual meeting of the American Educational Research Association, San Francisco.

Parsons, R.M. (2006), "External evaluation in New Zealand schooling", unpublished doctoral thesis, Massey University, Palmerston North.

Pekrul, S. and B. Levin (2007), "Building student voice for school improvement", in D. Thiessen and A. Cook-Sather (eds.), *International Handbook of Student Experience in Elementary and Secondary School*, Springer, Dordrecht, pp. 711–726.

Perryman, J., S. Ball, M. Maguire and A. Braun (2011), "Life in the pressure cooker – school league tables and English and mathematics teachers' responses to accountability in a results-driven era", *British Journal of Educational Studies*, Vol. 59(2), pp. 179-195.

Pont, B., D. Nusche and H. Moorman (2008), *Improving School Leadership – Volume 1: Policy and Practice*, OECD Publishing, Paris.

Portuguese Ministry of Education and Science (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Portugal*, *www.oecd.org/edu/evaluationpolicy*.

RBAG (2008), *Reducing Burdens Action Group – Report to Ministers*, September 2008.

Reynolds, D., R. Bollen, B. Creemers, D. Hopkins, L. Stoll and N. Lagerweij (1996), *Making Good Schools: Linking School Effectiveness and School Improvement*, Routledge, London.

Reynolds, D. and C. Teddlie (2000), "The processes of school effectiveness", in C. Teddlie and D. Reynolds (eds.), *The International Handbook of School Effectiveness Research*, Falmer Press, London.

Roald, K. (2010), *Kvalitetsvurdering som organisasjonslæring mellom skole og skoleeigar*, University of Bergen, Norway.

Robinson, V.M.J., C. Lloyd and K.J. Rowe (2008), "The impact of leadership on student outcomes: An analysis of the differential effects of leadership type", *Education Administration Quarterly*, Vol. 44(5), pp. 635-674.

Ronnberg, L. (2011), "Exploring the intersection of marketisation and central state control through Swedish national school inspection", *Education Inquiry*, Vol. 2, No. 4, pp. 689-707.

Rosenkvist, M.A. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, *oecd.org/edu/workingpapers*.

Rosenthal, L. (2004), "Do school inspections improve school quality? Ofsted inspections and school examination results in the UK", *Economics of Education Review,* No. 23, pp. 143–151.

Rudd, P. and D. Davies (2000), "Evaluating school self-evaluation", paper presented at the British Educational Research Association Conference, Cardiff University, 7-10 September.

Rudduck, J. (2007), "Student voice, student engagement, and school reform", in D. Thiessen and A. Cook-Sather (eds.), *International Handbook of Student Experience in Elementary and Secondary School*, Springer, Dordrecht, pp. 587-610.

Sammons, P., J. Hillman and P. Mortimore (1995), *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*, Institute of Education, London, and the Office for Standards in Education.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012b), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Schagen, S. and C. Wylie (2009), *School Resources, Culture and Connections*, New Zealand Council of Educational Research (NZCER), Wellington.

Scheerens, J., M. Ehren, P. Sleegers and R. de Leeuw (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Netherlands*, University of Twente, the Netherlands, *www.oecd.org/edu/evaluationpolicy*.

Scheerens, J., H.W.C.H. van Amelsvoort and C. Donoughue (1999), "Aspects of the organisational and political context of school evaluation in four European countries", *Studies in Educational Evaluation*, Vol. 25, pp. 79–108.

Schildkamp, K., L.T.M. Rekers-Mombarg and T.J. Harms (2012), "Student group differences in examination results and utilisation for policy and school development", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, Vol. 23(2), Routledge, pp. 229-255.

Seashore Louis, K., K. Leithwood, K.L. Wahlstrom and S.E. Anderson (2010), *Learning from Leadership Project: Investigating the Links to Improved Student Learning*, final report of research findings, commissioned by the Wallace Foundation, University of Minnesota, University of Toronto, *www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf*.

SEP and INEE (Secretariat of Public Education and National Institute for Educational Assessment and Evaluation) (2011), *Country Background Report for Mexico*, prepared for the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes, *www.oecd.org/edu/evaluationpolicy*.

Shaw, I., D.P. Newton, M. Aitkin and R. Darnell (2003), "Do OFSTED inspections of secondary education make a difference to GCSE results?", *British Educational Research Journal*, Vol. 29, No. 1, pp. 63-76.

Sheppard, B. (1995), *Implementing Change: A Success Story*, Morning Watch Archives, Memorial University of Newfoundland and Labrador, *www.mun.ca/educ/faculty/mwatch/archives1.htm.*

Shewbridge, C., M. Ehren, C. Tamassia and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: Luxembourg*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., M. Hulshof, D. Nusche and L. Stoll (2011a), *OECD Reviews of Evaluation and Assessment in Education: School Evaluation in the Flemish Community of Belgium*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011b), *OECD Reviews of Evaluation and Assessment in Education: Denmark*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., J. van Bruggen, D. Nusche and P. Wright (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Slovak Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

SICI (Standing International Conference of Inspectorates) (2003), *Effective School Self-Evaluation* (ESSE), SICI, Belfast.

Simons, H. (1987), *Getting to Know Schools in a Democracy: The Politics and Process of Evaluation*, Falmer Press, London.

Sims, D. (2008), "Strategic responses to school accountability measures: It's all in the timing", *Economics of Education Review*, Vol. 27, No. 1, pp. 58-68.

Smyth, J. (2007), "Toward the pedagogically engaged school: Listening to student voice as a positive response to disengagement and 'dropping out'?", in D. Thiessen and A. Cook-Sather (eds.), *International Handbook of Student Experience in Elementary and Secondary School*, Springer, Dordrecht, pp. 635-658.

Stoll, L., J. Halbert and L. Kaser (2011), "Deepening learning in school-to-school networks", in C. Day (ed.), *International Handbook on Teacher and School Development*, Routledge, London.

Swaffield, S. and J. MacBeath (2005), "School self-evaluation and the role of the critical friend", *Cambridge Journal of Education*, Vol. 35 (2), pp. 239-52.

Swedish Ministry of Education and Research (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Sweden*, Ministry of Education and Research, Stockholm, *www.oecd.org/edu/evaluationpolicy*.

Teddlie, C., B. Creemers, L. Kyriakides, D. Muijs and F. Yu (2006), "The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs", *Educational Research and Evaluation*, Vol. 12, pp. 561-582.

van Bruggen, J. (2009), *Inspectorate Profile: Scotland*, The Standing International Conference of Inspectorates (SICI), *www.sici-inspectorates.eu/en/87080-scotland*.

van Bruggen, J. (2008), *Inspectorate Profile: England*, The Standing International Conference of Inspectorates (SICI), *www.sici-inspectorates.eu/en/87092-england*.

van de Grift, W. (2009), "Reliability and validity in measuring the value added of schools", *School Effectiveness and School Improvement*, Vol. 20 (2), pp. 269–285.

van de Grift, W. (2007), "Quality of teaching in four European countries: A review of the literature and application of an assessment instrument", *Educational Research*, Vol. 49, No. 2, pp. 127-152.

van Petegem, P. and J. Vanhoof (2009), "Zelfevaluatie als motor voor schoolontwikkeling – Succesfactoren en valkuilen" (Self-evaluation as engine for school development – Success factors and stumbling blocks), Wolters Plantyn, Mechelen, Belgium.

Vanhoof, J. and P. Van Petegem (2005), "Kwaliteit van scholen (openbaar) vergelijken? Het zwart-wit denken voorbij" (Comparing school quality? Beyond black-or-white thinking), Wolters Plantyn, Mechelen.

Vanhoof, J., P. Van Petegem and S. De Maeyer (2009), "Attitudes towards school self-evaluation", *Studies in Educational Evaluation*, Vol. 35, pp. 21-28.

Vašťatková, J. and M. Prášilová (2010), "Možné podoby autoevaluace ve školách" (Possible forms of self-evaluation in schools), *Orbis Scholae*, Vol. 4, Issue 1, pp. 73-89.

Visscher, A.J. and R. Coe (2003), "School performance feedback systems: Conceptualisation, analysis and reflection", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, Vol. 14:3, Routledge, pp. 321-349.

Visscher, A.J., P. Wild, and A. Fung (2001), *Information Technology in Educational Management; Synthesis of Experience, Research and Future Perspectives on Computer Assisted School Information Systems*, Kluwer Academic Publishers, Dordrecht/Boston/London.

Weir, S. (2011), *A Report on the First Phase of the Evaluation of DEIS. Summary Report*, Educational Research Centre, Dublin, *www.erc.ie/?s=4*.

West A., P. Mattei and J. Roberts (2011), "Accountability and sanctions in English schools", *British Journal of Educational Studies*, Vol. 59(1), pp. 41-62.

Willms, J.D. (1997), "Parental choice and education policy", *CES Briefing No. 12,* Centre for Educational Sociology, University of Edinburgh.

Wolf, I.F. de and F.J.G. Janssens (2007), "Effects and side effects of inspections and accountability in education: An overview of empirical studies", *Oxford Review of Education*, Vol. 33, Issue 3, pp. 379-396.

Yeung, S. (2011), "A school evaluation policy with a dual character: Evaluating the school evaluation policy in Hong Kong from the perspective of curriculum leaders", *Educational Management Administration and Leadership 2012*, Vol. 40, Sage, pp. 37-68.

Chapter 7

The appraisal of school leaders:

Fostering pedagogical leadership in schools

Backed by a growing research base, policy makers have increasingly recognised the significance of school leadership for effective teaching and learning. Within that context, an increasing number of countries have developed initiatives to strengthen the leadership capacity of their schools. While research on the effects of different appraisal schemes is limited, some evidence suggests potential benefits of the appraisal of individual school leaders as a means to communicate a vision of effective leadership and a tool to influence and improve school leaders' practices and behaviours. This chapter discusses the approaches that countries take to appraise individual school leaders. Building on a conceptualisation of school leadership and a discussion of drivers and contextual developments, it analyses the governance of appraisal schemes, appraisal procedures, the capacity required for effective appraisal, and the use of appraisal results. The chapter concludes with a set of pointers for future policy development.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

School leadership, as an individual as well as an organisational quality, plays a crucial role for both enhancing teaching and learning in schools as well as for building effective evaluation and assessment frameworks (see Chapter 3 for a conceptualisation of a holistic approach to evaluation and assessment; see Chapters 5 and 6 for the role of school leaders for teacher appraisal and school evaluation). Given the significance of strong leadership in schools, this chapter analyses countries' current approaches to the appraisal of individual school leaders and the ways in which appraisal can contribute to improving teaching and learning in schools. Although the evidence base on the effects of different models for the appraisal of individual school leaders is limited, and appraisal is always only one of many influences on school leaders' practices and behaviours, a few studies have pointed towards the potential benefits of appraisal as a tool for improving school leadership and an opportunity for professional feedback (also see Radinger, forthcoming, for a literature review of school leader appraisal).

The chapter and analysis of school leader appraisal is based on a conceptualisation of school leadership. While school leadership styles are not exclusive, and context is essential, the chapter highlights the overall role of pedagogical leadership for improved teaching and learning. It also points out the need to consider the impact of appraisal on new and innovative models of school leadership, such as distributed leadership and system leadership. The chapter seeks to pay due attention to drivers and contextual developments that influence considerations about school leader appraisal (e.g. greater decentralisation and school autonomy, changing school leadership roles, and challenges facing the school leadership profession). Following the analytical approach of the overall report, the chapter, then, analyses the governance of appraisal systems, the procedures that are applied, the capacity required for effective appraisal, and the ways in which results are used. Based on this discussion of current practices as well as related research evidence, a number of current policy challenges are identified throughout the chapter. The chapter concludes with a set of pointers for future policy development for countries to consider. It is, however, important to bear in mind that the relevance of both policy challenges and policy pointers is highly context-specific and that the particular approach to school leader appraisal depends on the system concerned.

Analytical approach

Prior to the in-depth analysis of approaches to school leader appraisal the chapter examines different conceptualisations of school leadership. It sketches the scope of the discussion and develops a conceptual framework for the subsequent analysis.

Conceptualising school leadership

Leadership, in general, and leadership in schools is difficult to define. There is little consensus about what leadership is, how it emerges and how it relates to concepts like management and administration. Traditionally, debate has focussed on individual school leaders and the question as to what extent leadership stems from an individual's personal qualities, traits and characteristics (i.e. who a leader is) or from an individual's actions and behaviours (i.e. what a leader does) (Bush and Glover, 2003; Krüger and Scheerens, 2012). Recently, leadership in schools has been described as a process of intentional influence on activities and relationships that is based on a clear sense of direction (Bush and Glover, 2003; Pont et al., 2008a; Louis et al., 2010). Within this conceptualisation of

the nature of leadership, one line of research has sought to establish the core areas of schools that school leaders need to influence to exercise effective leadership.[1]

The nature of effective school leadership

While various paradigms of effective school leadership have emerged through time (e.g. instructional, transformational and transactional leadership models), research has recently stressed the importance of pedagogical leadership, i.e. leadership that is focused on the improvement of teaching and learning (Bush and Glover, 2003; Mulford, 2008; Krüger and Scheerens, 2012).[2] This vision is closely linked to a vision of schools as professional communities (Spillane and Kenney, 2012). Based on the premise that school leaders need to direct their efforts to the improvement of teaching and learning, research has identified a set of leadership practices and behaviours that form the basis of pedagogical leadership in many contexts. It is, however, important to bear in mind that, while research has stressed the overall benefits of pedagogical leadership styles, different leadership styles are not exclusive and tasks that involve building organisational routines and that are more administrative and management-oriented are, at times, equally important (Krüger and Scheerens, 2012). The wide range of tasks and responsibilities that school leaders are often expected to fulfil also bear a risk of placing too high expectation on school leaders that are romanticised in a heroic way (Pont et al., 2008a; Coffield, 2012).

Day et al. (2009, 2010) provided an updated version of the commonly cited core leadership practices of setting directions, redesigning the organisation, developing staff and managing teaching and learning, identified in a literature review by Leithwood et al. (2006). Day et al. (2010) conceptualised successful school leaders as leaders that:

- define their values and vision to raise expectations, set directions and build trust
- reshape the conditions for teaching and learning
- restructure parts of the organisation and redesign leadership roles and responsibilities
- enrich the curriculum
- enhance teacher quality
- enhance the quality of teaching and learning
- build collaboration internally
- build strong relationships outside the school community.

Pont et al. (2008a) provided a similar definition of successful school leadership for improved student outcomes that draws on the following set of interrelated practices and behaviours:

- leadership focused on supporting, evaluating and developing teacher quality
- setting learning objectives and implementing intelligent assessment systems
- strategic use of resources and their alignment with pedagogical purposes
- school leadership beyond the school borders.

However, there has been a growing realisation that successful school leadership is always context-dependent and that one size of leadership may not necessarily fit all

circumstances. The school level, the size of a school, the governance of a school, the staffing of a school, the socio-economic background of students, the level of parental involvement, as well as local, regional and country contexts all influence what works as effective school leadership (Bush and Glover, 2003). Notions of situational and contingent leadership have stressed this role of organisational factors and structures that influence the kind of leadership that is required (Spillane et al., 2004) To give an example for the role of these contextual factors, primary school principals are often generalists, secondary school principals often subject specialists, which influences the degree of subject-specific support principals can give to teachers. In primary schools as well as smaller schools in rural areas, school principals often have to work with fewer resources and lack administrative assistance and senior staff who could support them. Often responsible for teaching and school management, administration and leadership, many school principals in primary and small rural schools, therefore, face role overload (Pont et al., 2008a; Clifford and Ross, 2011). At the same time, particular contexts also influence the impact and perceptions that different leadership practices and behaviours have in schools (Hallinger and Heck, 1998; Louis et al., 2010). Louis et al. (2010), for example, found that effective leadership is usually perceived differently depending on school size and type. In their study, teachers in primary or small schools often experienced leadership as more effective than in secondary or large schools. Concerning their sets of core leadership practices and behaviours, both Day et al. (2010) and Pont et al. (2008a), therefore, stressed that different contexts require a different leadership response. Contexts, then, influence how leaders exercise similar core practices of successful leadership. School leaders' awareness of the contexts in which they work and the ways in which they adjust their leadership to these contexts forms a fundamental part of effective school leadership (Leithwood et al., 2004, Louis et al., 2010).

These research insights into the nature of effective school leadership, the overall benefits of pedagogical leadership styles for the improvement of teaching and learning, the highly contextual nature of school leadership and the danger of heroic visions of school leadership present challenges for school leaders and influence approaches to school leader appraisal. Considering the role of appraisal in communicating school leaders' expected tasks and responsibilities and related competences (Catano and Stronge, 2007), policy makers face the challenge of taking research on effective school leadership in appraisal frameworks into account. Policy makers, however, also need to find ways to minimise risks that appraisal can also lead to role overload and increased levels of stress among school leaders (Normore, 2004; Pashiardis and Brauckman, 2008). Furthermore, in light of the highly contextual nature of effective school leadership, appraisal systems need to ensure flexibility and adaptability to respond to distinctive school and community contexts and to the circumstances school leaders face in these particular conditions (Leon et al., 2011).

The development of new and diverse models of school leadership

Traditionally, leadership in schools has referred to individual formal leadership roles, such as the school principal, deputy school principal, middle leaders and department heads. Without assuming that everyone is or should be a leader and without negating the role of individual leaders, distributed perspectives[3] on school leadership have increasingly recognised that school leadership does not necessarily reside in a formal position or the authority of a single person (see, for example, Bennet et al., 2003, for a literature review of distributed school leadership). Rather, leadership can be practiced by different actors and be seen as an organisational quality. Leadership, if defined as a process of intentional

influence on activities and relationships that is based on goals and a sense of direction, can be practiced by a range of actors within schools through both formally designated positions and more informal ways (Bush and Glover, 2003; Pont et al., 2008a). Conceptualised as an organisational quality, leadership stretches across individuals as a social practice. Spillane et al. (2004), for instance, defined leadership as a practice that emerges out of the interaction of individuals, i.e. leaders and followers, within the context of a specific situation (see Box 7.1 for an appraisal approach in the Northern Territory in Australia that focuses on school leadership as an organisational quality and closely aligns the school evaluation and school leader appraisal processes).

Box 7.1 Aligning the school evaluation process and the individual appraisal of school leaders: The School Accountability and Performance Improvement Framework in the Northern Territory, Australia

In 2006, the Northern Territory introduced a School Accountability and Performance Improvement Framework as an accountability system that is designed to set clear performance standards and to promote a culture of evidence-based decision making and continuous improvement in the Northern Territory's schools. Based on the premise of a strong interrelationship between the quality of school leadership and the quality of teaching and student learning, the framework integrates external and school self-evaluation processes, individual performance plans for all school staff and school principal performance review and appraisal. As the Department of Education and Training of the Northern Territory states, the framework is built around a vision of effective leadership "that engages people at all levels of the organisation in the learning process by creating a culture of inquiry which develops new capabilities and revolutionises teaching and learning" as the basis for school improvement.

The framework is directed towards three focus areas: 1) a quality educational experience for all students; 2) responsiveness to students, parents and the community; and 3) strong school leadership and vision. Within these overall objectives, it is closely aligned with local and system performance goals and targets, including those set by the Northern Territory and Australian governments (e.g. through the requirement that the development of priorities, strategies and targets based on the School Accountability and Performance Improvement Framework needs to reflect funding arrangements with Australian and/or Northern Territory authorities).

At the core of the School Accountability and Performance Improvement Framework lies a **school self-evaluation** process that is coupled with an **external school evaluation** procedure through an external evaluator, the Director School Performance (DSP). Both of these processes are structured around the formulation of a 4-year **Strategic Improvement Plan** that identifies key priorities, strategies and targets and that is subject to ongoing review. The Strategic Improvement Plan must be framed around a set of 5 **Key Result Areas**: 1) teaching and learning; 2) well-being; 3) participation, transitions and pathways; 4) partnerships; and 5) leadership. The Key Result Area of leadership is defined through a range of **outputs/milestones** (includes school mission/vision and values statements; 360 degree feedback; interpersonal relationships; school review; performance management processes; professional development data; leadership and professional development components of School Literacy and Numeracy and ICT self-assessments; alignment of expenditure with strategic directions; and expenditure of supplementary funding within period of agreement) and **outcomes/targets** (includes occupational health and safety data and staff turnover).

Box 7.1 Aligning the school evaluation process and the individual appraisal of school leaders: The School Accountability and Performance Improvement Framework in the Northern Territory, Australia (*continued*)

An **Annual Operational Plan** defines the school's short-term priorities to achieve and implement the wider goals that are laid out in the 4-year Strategic Improvement Plan. The monitoring of effective financial management practices constitutes an integral element of the Annual Operational Planning process. Schools must demonstrate their progress towards the Annual Operational Plan in the form of a **School Annual Performance Report** that is subject to the endorsement by the DSP and the school council.

In alignment with the Strategic Improvement Plan, all **school staff** are required to develop individual **Performance Plans**. This process, which is managed by school principals, intends to build capacity and leadership across the school, focus the attention of staff on teaching and learning and align resources with professional development needs. **School principals** themselves are required to engage in an **annual performance review** and **appraisal process**. The annual performance review builds on a school's performance information and evidence for compliance with legal requirements and targets that are set by the authorities of the Northern Territory and the Australian government. The school principal performance review and appraisal process aims to improve school leader's practices and behaviours, but also informs school principals' contract renewal process. It is aligned with the School Accountability and Improvement Framework through the Key Results Areas that include leadership-specific responsibilities, the results of the school review and clear guidelines on effective leadership.

Source: Northern Territory Government, Department of Education and Children's Services website, Accountability and Performance Improvement Framework (APIF), see *www.det.nt.gov.au/about-us/APIF*.

While leadership thus conceptualised is inherently distributed within schools, leadership can also be distributed within schools in more formal and co-ordinated ways (e.g. through opportunities for teacher leadership and leadership teams and committees) (Pont et al., 2008a; Schleicher, 2012) (see Box 7.2 for an appraisal approach in Chile explicitly designed to encourage and improve the distribution of leadership through leadership teams). However, formal arrangements for distributed leadership are still rare. While research suggests that a distribution of leadership tasks and responsibilities across individuals, including teachers, and teams may contribute to greater overall leadership capacity in schools, help foster change, and sustain that improvement over time (Mulford, 2003; Pont et al., 2008a), research on the effective distribution of leadership and the ways in which different arrangements influence school outcomes is still scarce (Harris and Spillane, 2008). It is also important to bear in mind that the distribution of leadership depends on particular contexts (e.g. school size) and that it creates its own challenges at the same time (Louis et al., 2010). As Mulford (2008) highlighted, successful teacher leadership, for example, depends on school leaders that create the conditions for effective teacher leadership to emerge (e.g. through building structures for ongoing collaboration), which, in turn, requires ongoing support for school leaders.

Box 7.2 The appraisal of distributed school leadership: The Appraisal of Collective Performance process (*Asignación de Desempeño Colectivo*) in Chile

Based on the belief that leadership is an organisational quality, the Appraisal of Collective Performance process (*Asignación de Desempeño Colectivo*) constitutes an appraisal process designed to encourage school leaders to collaborate and improve the practices and behaviours of leadership teams. Through the setting and evaluation of collective objectives and targets, the system aims to strengthen the organisational leadership and management capacity of schools as a whole, to encourage school leaders' commitment to the improvement of teaching and learning in their school, and to facilitate and encourage successful collaboration of leadership teams Participation in the Appraisal of Collective Performance process is voluntary for school leaders in all schools with at least 250 students. First implemented in 2005, an increasing number of school leadership teams are taking part. The Appraisal of Collective Performance process has been complemented with a wide range of further policies designed to strengthen the leadership capacity of Chilean schools, such as an individual Performance Appraisal system (*Evaluación de Desempeño*) implemented in 2006, and further measures linked to the selection and recruitment of principals, principals' responsibilities and professional development introduced through the Quality and Equality of Education Law (Law No. 20.501) in January 2011.

The Appraisal of Collective Performance system is built around the development of an institutional objective (*objetivo institucional*) and two to four institutional targets (*metas institucionales*) that results in an agreement of collaborative leadership (*Convenio de desempeño colectivo*). The development of the agreement of collaborative leadership involves the distribution of related tasks and school principals and other technical-pedagogical school leaders taking on mutual responsibilities. The institutional targets are determined by local contexts and priorities, but need to be linked to five areas of school leadership (leadership, pedagogical leadership, school climate, support to students, financial management) and results. At least one of the institutional targets needs to be related to pedagogical leadership, another to results. To align the Appraisal of Collective Performance process with other measures for improving school leadership, the institutional objective and institutional targets need to be defined with reference to the annual municipal education development plan (*Plan Anual de Desarrollo de la Educación Municipal* [PADEM]) and the school development plan (*Proyecto Educativo Institucional* [PEI]). For school leadership teams in public schools, the institutional targets of the voluntary Appraisal of Collective Performance process can be tied to the mandatory individual school leader appraisal process (*Evaluación de Desempeño de Docentes Directivos y Técnico-Pedagógicos*) to create further alignment and synergies.

Once a school leadership team has established an agreement of collaborative leadership, the school leadership team together with its school organising body develop strategies to monitor the achievement of the objective and targets and related leadership practices and behaviours. It is, then, subject to further approval or revision by the responsible provincial department of education (*Departamento Provincial de Educación*) to ensure the agreement complies with legislation and corresponds to national education goals. Ultimately, the school leadership team presents its agreement, institutional objective and targets to the school community.

The extent to which the institutional objective and each of the institutional targets have been reached is assessed by the school organising body on the basis of an implementation report (*Reporte de Implementación*) and an evidence portfolio (*Carpeta de Evidencias*) compiled by the school leadership team to document progress towards the objective and targets and any challenges and difficulties in the process. The extent to which the overall agreement of collaborative leadership has been reached is calculated from the total sum of the degree of achievement of the individual objective and targets. The results are, ultimately passed on to the provincial department of education for validation. On the basis of four rating levels (0-49.9%, 50-74.9%, 75-89.9%, 90-100%), school leadership teams that have achieved the two highest rating categories receive a financial bonus of 10% and 20% of the national basic minimum remuneration (*Remuneración Básica Mínima Nacional* [RBMN]). To ensure the sound implementation of this tool, the Chilean Ministry of Education can carry out external audits of the Appraisal of Collective Performance process as implemented by individual school organising bodies.

Source: Centre for Pedagogical Training, Experimentation and Research (CPEIP) website on Appraisal of Collective Performance process, see *www2.gestionyliderazgoeducativo.cl/gestioncalidad/asignacion/home/index.php*.

At the same time, new formal school leadership roles have emerged in various contexts that stress the importance of system leadership and school leaders' involvement in other schools (see, for example, Pont et al., 2008b for a series of case studies on system leadership). From a system-wide perspective school leaders take responsibility for and work towards the success of other schools as well as their own in order to foster improvement across the education system (Hopkins, 2008). Mulford (2008) described this vision of school leadership that includes attention to the bigger picture through the notion of sustainable leadership. To give two examples for current practices, in England, a variety of system leadership roles have emerged. These include both structured roles such as consultant leaders, school improvement partners and national leaders of education as well as more informal roles that emerge locally (PricewaterhouseCoopers, 2007; Hopkins, 2008). In Finland, school leaders in some municipalities also act as district leaders devoting two-thirds of their time to their own schools and one-third to the district (Pont et al., 2008a). However, as in the case of distributed leadership, such practices are still rather rare and evidence on the effects of different approaches is limited.

The conceptual developments towards new and diverse forms of leadership, such as distributed leadership and system leadership, require policy makers to take the effects of appraisal on innovative leadership concepts, distributed and system leadership, into account. Accountability requirements may pose one barrier for encouraging school leaders to distribute tasks and responsibilities, for example (Mulford, 2008). At the same time, appraisal procedures need to reflect the emerging demands that come with changing leadership roles.

Scope and definitions

Based on these conceptions of school leadership, this section provides a set of working definitions. It also lays out the scope of the chapter in relation to the overall evaluation and assessment framework and teacher appraisal and school evaluation processes.

Defining school leader appraisal

This chapter explores approaches to the *appraisal* of school leaders defined as procedures involving external evaluators that seek to identify school leaders' strengths and weaknesses using objective criteria in order to make a judgement about their competences for *performance management*, *employment-related decisions* and/or *rewards purposes*. Appraisal schemes linked to the selection and recruitment of school leaders or to the completion of school leaders' probationary period are not covered in this chapter. The chapter pays particular attention to policy initiatives in countries participating in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes that have established central/state frameworks for the appraisal of school leaders. In various contexts, the requirements for appraisal, responsibilities and/or procedures differ between leadership positions (e.g. school principals and deputy principals), school levels (e.g. primary and secondary education), or school types (e.g. public and private schools). Such differences and the particular context which research evidence refers to are clearly indicated throughout the chapter.

Despite developments towards more diverse conceptions of school leadership, such as distributed leadership and system leadership, appraisal in many contexts focuses predominantly on individual formal school leadership positions (see Table 7.A.1 for an overview of practices across countries). Therefore, this chapter focuses on approaches to the appraisal of *individual school leaders* that have already taken up their position.[4] Reflecting

current practices, the terms *school leadership* and *school leaders* are used interchangeably in this chapter to refer to the roles and dispositions of individuals, in particular to the highest leadership positions in a school. i.e. school principals and deputy principals.[5]

School leader appraisal in relation to teacher appraisal and school evaluation

It is also important to define the scope of this chapter and the nature of school leader appraisal processes in relation to further elements of the evaluation and assessment framework, and teacher appraisal and school evaluation, in particular (also see Chapter 3). While some systems base their school leader appraisal systems on their regulations for teacher appraisal, others also evaluate the quality of leadership in their schools through school evaluation processes.

Systems which base their procedures for the appraisal of school leaders on their policy framework for teacher appraisal include, among others, Belgium (Flemish Community), France (ISCED level 1), Korea (School Principal Appraisal for Professional Development), Mexico (schools at ISCED levels 1 and 2 managed by state level education authorities or autonomous agencies),[6] Poland, Slovenia and Northern Ireland in the United Kingdom. Various other countries evaluate the quality of school leadership in a school solely through external school evaluations (e.g. Austria) or in addition to individual appraisal processes (e.g. the Czech Republic, New Zealand, Poland, Portugal, Sweden and Northern Ireland in the United Kingdom). To give an example, in Northern Ireland, school principals are subject to an individual annual Performance Review and Staff Development Scheme (PRSD), which is based on the teacher appraisal process. In addition, the Education and Training Inspectorate (ETI) evaluates schools against a set of quality indicators as set out in the reference document *Together Towards Improvement*. This process, which seeks to promote high-quality learning and teaching and to provide information about the quality of education to the Northern Ireland Department of Education, also includes an evaluation of the effectiveness of the school's organisation, leadership and management. (Department of Education, Northern Ireland, forthcoming). Individual school leader appraisal schemes that are based on teacher appraisal processes are discussed as part of this chapter to the extent that they relate to the appraisal of school leaders. The evaluation of a school's leadership capacity as an organisational quality within the framework of school evaluations is covered as part of Chapter 6.

The potential coexistence of different approaches to the evaluation of a school's leadership capacity requires a reflection about the ways in which different processes, and school evaluation and school leader appraisal, in particular, inform each other. One of the challenges identified by the OECD Country Review of New Zealand, for example, concerns the articulation of the individual school leader appraisal process conducted by individual school boards of trustees and the evaluation of school leadership through the external school review process carried out by the Education Review Office. This includes challenges concerning the systematic sharing of information and the aspects and criteria used to describe effective leadership in both processes (Nusche et al., 2012). Poland and Portugal provide two interesting examples for ways in which to link school evaluation and school leader appraisal processes. In Poland, school evaluation and school leader appraisal processes are aligned through the requirement that the evaluators of individual school leaders need to take the results of school evaluations into account when carrying out an individual appraisal. In Portugal, the appraisal of school principals in public schools in Portugal consists of two separate processes. School leaders are appraised individually by a general council. In addition, school leaders are evaluated by the Inspectorate as part of the school evaluation process. The results of both processes are taken into account to different

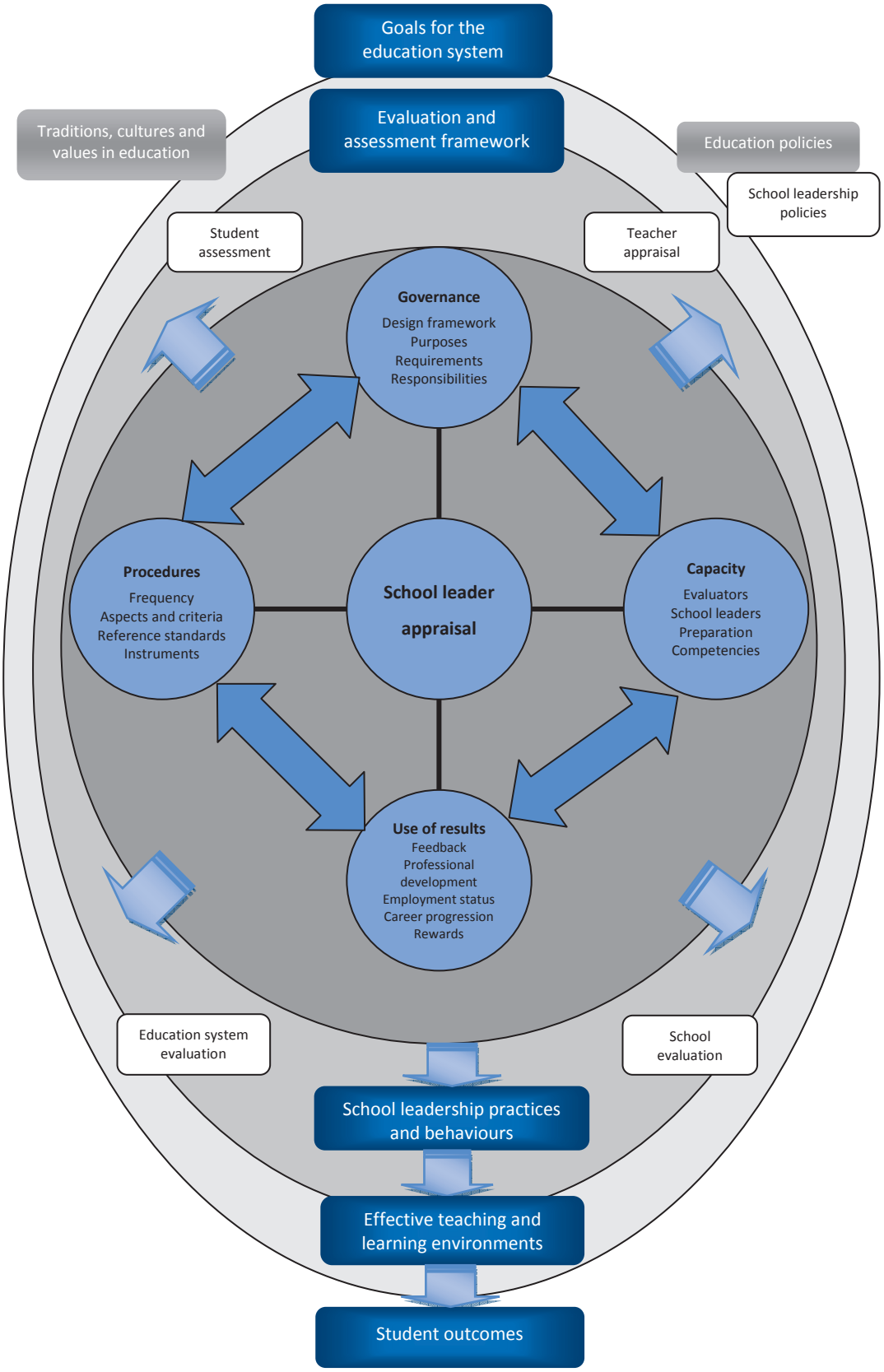
degrees. The result of the individual appraisal counts for 60%, the result of the school evaluation process for 40%.[7] The possibility to evaluate the performance of school leaders through school evaluation processes, teacher appraisal schemes or a separate procedure that is exclusively designed for the appraisal of individual school leaders also raises questions about the relative merits, benefits and disadvantages of each of these approaches.

A conceptual framework

The following provides a conceptual framework summarising the various aspects involved in school leader appraisal and the ways in which these interconnect (see Figure 7.1). Starting from the assumption that the overarching objective of school leader appraisal lies in the improvement of school leadership practices and behaviours for improved teaching and learning, and embedded within the wider evaluation and assessment framework (also see Chapter 3), this chapter discusses the following four elements of appraisal: (1) governance, (2) procedures, (3) capacity, and (4) use of results.

- **Governance**: This aspect concerns the overall design and organisation of school leader appraisal. This includes the setting of appraisal requirements and the distribution of responsibilities for the design of school leader appraisal. It also includes the objectives of a particular school leader appraisal system. These are framed within the goals for a particular education system and the overall purpose of improving school leadership practices and behaviours for improved teaching and learning.
- **Procedures**: This aspect refers to the features of school leader appraisal and the ways in which these are combined to create a specific appraisal model. This includes requirements for the frequency of appraisal, the definition of appraisal aspects and criteria, the use of reference standards, the alignment with professional leadership standards and the combination of instruments to gather relevant information. To give an example, a legal framework for appraisal may require school leaders and evaluators to develop individual performance objectives that are informed by a set of professional school leadership standards, individual development needs, the school development plan and wider system needs. The achievement of the performance objectives may subsequently be evaluated through meetings between an evaluator team and the school leader and sources of information chosen collaboratively by the evaluator team and the school leader. These might include school visits, classroom observations, a leadership portfolio and teacher and student questionnaires.
- **Capacity**: This aspect analyses the distribution of responsibilities for the implementation of appraisal and the preparation to appraise, to be appraised and to use the results of an appraisal. It includes questions about the necessary competences of both evaluators and school leaders to perform an appraisal and the development of the capacity to use the results of an appraisal for the improvement of school leaders' practices and behaviours.
- **Use of results**: This aspect concerns the mechanisms that ensure that appraisal results are used so that the objectives of an appraisal are reached. In line with the objectives of appraisal, appraisal results can be used in different ways. These include giving performance feedback, informing professional development plans and informing decisions about a school leader's employment status, career advancement and/or financial and other rewards.

Figure 7.1 Conceptual framework for school leadership appraisal



Impact, drivers and contextual developments

When considering approaches to the appraisal of school leaders, it is essential to bear in mind school leaders' role for teaching and learning as well as for the implementation of education policies. It is also crucial to take drivers and contextual developments that affect the school leadership profession into account.

The importance and impact of school leader appraisal

Education systems have increasingly recognised the importance of school leadership for improving teaching and learning as well as for ensuring that education policies reach the classroom, as substantiated in a recent OECD project on Improving School Leadership (Pont et al., 2008a). Subsequent research has further strengthened arguments that school leaders that take direct responsibility for pedagogical leadership and for the quality of education in their schools can make a strong difference to teaching and learning. Research has also highlighted the key role of effective school leadership for the successful implementation of policies, including in the field of evaluation and assessment. Research evidence for the positive impact of successful leadership on schools and the implementation of policies bolsters arguments for the effective recruitment, career development, appraisal and retention of school leaders. While research on the effects of different appraisal schemes is rather limited, some evidence suggests potential benefits of appraisal as a means to communicate a vision of effective leadership, such as pedagogical leadership, and as a tool to influence and improve school leaders' practices and behaviours.

School leadership can have a strong influence on the creation of effective teaching and learning environments and indirectly influence student outcomes

Since the 1970s education research has explored the role that school leaders play for student learning. An article by Hallinger and Heck (1998) reviewed quantitative research conducted in different national contexts between 1980 and 1995 that produced mixed findings on the relationship between school principals' beliefs and behaviours and student achievement. While acknowledging the need for further research in this area and the limitations of conducting research on a constantly evolving concept, Hallinger and Heck suggested a relatively small, but statistically significant effect of school leadership on school effectiveness and improvement. Employing a theoretical framework of direct, mediated and reciprocal effects, and in light of a lack of data on reciprocal effect studies, their review indicated that most school leader effects on student outcomes are indirect, i.e. mediated through other people, events and organisational factors that are internal and external to a school. Marzano et al. (2005) conducted a quantitative meta-analysis of research on school principals' leadership from 1978 to 2001 undertaken in the United States. Bearing in mind that this study specifically analysed the effects of school principals in the United States and did not distinguish between different channels of influence, Marzano et al.'s meta-analysis went further than Hallinger and Heck arguing that school principals can have a strong effect on student achievement.

Albeit some scholars have cautioned about policy running ahead of the evidence on the importance of school leadership (e.g. Coffield, 2012), a series of recent research projects has further strengthened the evidence base for a relationship between successful school leadership, effective teaching and learning, and improved student outcomes that

previous research had already pointed to. Even though a number of conceptual and methodological issues remain (see, for example, Pont et al., 2008a), research has established strong indications for the positive, mainly indirect, effects of successful school leadership on student outcomes and the ways in which school leaders can influence conditions for teaching and learning. Research in the United States has investigated the relationship between school leadership and student achievement (Leithwood et al., 2004, Louis et al., 2010).[8] Examining school leadership from a broader perspective considering leadership at the state, district and school levels as well as the interrelation between these dimensions, Leithwood et al. and Louis et al. argued that school leadership is second only to classroom instruction among school-related factors that influence what students learn in schools. They also found that the positive relationship between school leadership and student achievement is mainly an indirect one that rests on leaders building synergies between factors that influence student learning. Research in England has equally strengthened arguments for an influence of school leadership, with particular reference to the leadership of headteachers, on student outcomes and illuminated the ways in which school leaders influence these outcomes (Leithwood et al., 2006; Day et al., 2009, 2010).[9] As part of an OECD project on Improving School Leadership across 22 OECD countries, Pont et al. (2008a) reviewed research on the relationship between school leadership and student outcomes. Taking a broad view on school leadership that included school principals, deputy principals, leadership teams, school governing boards and other school-level professional personnel, Pont et al. also stressed "the pivotal role of school leadership in making schools more effective" through their influence on the motivations, capacities and working conditions of teachers that, then, shape classroom practices and student learning.

Furthermore, research has produced increasing evidence that effective school leadership can play a particularly important role in low-performing and disadvantaged schools. The research project by Leithwood et al. (2004) and Louis et al. (2010), for instance, highlighted that the relationship between school leadership and student achievement tends to be even stronger in low-performing schools. Based on such evidence, a recent OECD report on equity and quality in education identified support for school leaders (e.g. through training programmes, coaching and mentoring) as one of five policy levers for improving low-performing disadvantaged schools (OECD, 2012).

Improving school leadership can constitute a cost-efficient lever for the successful implementation of education policies and reforms as well as of evaluation and assessment frameworks

School leaders not only function as a strong source of influence and direction within schools and as a key intermediary between the classroom and the school as a whole, but also as a key channel between classrooms, schools and the entire education system, i.e. between policy and practice (Hopkins, 2008). The success of a new education policy depends to a great part on the ways in which school leaders promote the adaptation of school processes, cultures, attitudes and behaviours to a changing external context and how they engage teachers and students in this process (Pont et al., 2008a). Leithwood et al. (2004) made a similar point stressing that the success of a policy initiative always also depends on the ways in which school leaders integrate external and local improvement efforts, the extent of their support for teachers to adapt their practices, and their efforts to establish trust and collaboration with local stakeholders. Spillane and Kenney (2012) stressed the importance of school leaders that develop new formal organisational structures and routines when discussing ways in which school leaders

reconcile demands for the external legitimacy and for the safeguarding of the internal integrity of their school in light of a changing context towards accountability in education in the United States. As they argued, school leaders play a pivotal mediating role between an increasingly demanding external environment on issues of teaching, on the one hand, and teachers that value professional autonomy, on the other hand.

The effective operation of evaluation and assessment frameworks equally depends to a great extent on successful leadership in schools. In various contexts, school leaders carry key responsibilities for appraising the teachers of their school as part of the respective performance management processes (see Chapter 5). As Chapter 5 pointed out, school leaders can play an essential role in teacher appraisal processes to identify areas for professional development of individual teachers, and for preparing individual development plans that take the overall school development into account. School leaders are also key actors for implementing school self-evaluation and for collaborating in the external evaluation of their school in many contexts (see Chapter 6). As Chapter 6 highlighted, effective school self-evaluation depends to a large degree on school leaders' ability to stimulate engagement, to mobilise resources, and to ensure appropriate training and support. Research by Emstad (2011) examined the role of school principals for the successful implementation of a formative self-evaluation process in two Norwegian primary schools.[10] Her case studies illustrate the importance of school leadership for establishing favourable conditions for self-evaluation and for integrating the process into the daily operation of the school (e.g. through establishing shared goals and gaining support among teachers). Depending on the school principal's leadership, the self-evaluation processes were perceived differently by teachers in both schools in terms of importance, time pressures and the self-evaluation's value for improving teaching and for informing decision making at a school and an individual level. Based on these insights, it is difficult to envisage effective teacher appraisal, school self-evaluation or external school evaluation without strong leadership capacity (see Chapters 3, 5 and 6). Considering the importance of effective school leadership for the overall evaluation and assessment framework, there are, then, strong arguments for establishing adequate arrangements for the effective appraisal of school leaders.

Leithwood et al. (2004) furthermore, pointed out that, considering the potential impact of a relatively small group of school leaders on the implementation of education policies, such as policies related to evaluation and assessment, efforts to improve school leader recruitment, ongoing professional development, career advancement and appraisal can constitute highly cost-effective measures for making education policies effective and for improving teaching and learning for all students. The 2012 International Summit on the Teaching Profession highlighted the interest in improving school leadership across various countries, partly stemming from the high rates of return for investments into the improvement of a small, but central group in schools (Asia Society, 2012; Schleicher, 2012).

The appraisal of school leaders may constitute one element of a comprehensive set of policies to foster successful leadership in schools

While research on the effects of school leader appraisal is limited (Parylo, 2012a), a small number of recent studies suggest that appraisal, depending on the way it is designed and implemented, can constitute one tool to effectively influence school leaders' practices and behaviours. Sun and Youngs (2009)[11] and Sun et al. (2012)[12] examined the relationship between school principal appraisal and school principals' pedagogical leadership practices and behaviours in Michigan, United States, and Beijing, China. The

results of both studies, even though not meant to draw causal inferences between appraisal and leaders' practices and behaviours, suggest that appraisal can constitute one effective way for communicating a vision of effective school leadership and for influencing school leaders' practices and behaviours. This research, however, also highlighted the need to embed appraisal within wider support mechanisms to foster successful school leadership that focuses on the improvement of teaching and learning (e.g. including the effective recruitment of highly qualified candidates). Parylo et al. (2012b)[13] examined the experiences and perceptions of appraisal of a sample of school principals in four school districts in Georgia, United States. Even though this study revealed the potential for tensions in the appraisal of school leaders, participants overall experienced appraisal as a support system that provided an opportunity for reflection and professional growth and helped increase school leaders' awareness of their role for the improvement of teaching and learning.

However, Davis and Hensley (1999) as well as Ginsberg and Berry (1990) stressed the common reluctance of people in general to be subject to an evaluation, which already highlights the need to create appraisal systems that school leaders themselves perceive as helpful and useful. A few studies that have examined the perception and response of school leaders and evaluators to appraisal also pointed out that school leaders may not always perceive appraisal as a meaningful exercise (Davis and Hensley, 1999; Thomas et al., 2000; Gaziel, 2008). Davis and Hensley's study, for example, pointed towards the potentially political dimension of appraisal. Research cited by Parylo et al. (2012b) highlighted the risk that appraisal that focuses too much on a school leader's interpersonal relationships may contribute to higher stress levels for school leaders. Further research has highlighted the burgeoning workload many school leaders face in various countries (Pont et al., 2008a). In England in the United Kingdom, for example, a study by PriceWaterhouseCoopers (2007) revealed that 61% of the headteachers that took part in the study described their work-life balance as poor or very poor. While a survey of newly appointed principals carried out by PriceWaterhouseCoopers (2009) in Northern Ireland in the United Kingdom suggested that newly appointed principals are largely happy in their role, 40% of the principals that took part in a survey in the North of Ireland and 58% of principals that took part in a survey in the South of Ireland described their work-life balance as poor or very poor. In light of these risks, policy makers face the challenge of developing school leader appraisal processes, frameworks and conditions that do not require an excessive investment of time and efforts, that serve as an effective tool for improving school leaders' practices and behaviours and that are accepted and experienced as such by school leaders themselves.

Drivers and contextual influences

Various drivers and contextual developments have a bearing on general considerations about appraisal, whether to implement it and if so, how it should be designed. These include the growing importance of school leadership as illustrated through recent policy initiatives to strengthen school leaders' development, a trend towards greater decentralisation and school autonomy resulting in changing leadership roles, an awareness of the shortcomings of current appraisal approaches in various contexts and a growing policy concern about the challenges that school leaders face.

Comprehensive approaches to school leadership development are growing in importance

As the OECD Improving School Leadership project (Pont et al., 2008a) and the 2012 International Summit on the Teaching Profession (Asia Society, 2012; Schleicher, 2012) illustrated, effective school leadership is on the agenda of many countries. To provide a few examples, in Scotland, the education agenda Ambitious, Excellent Schools introduced in 2004 recognised the importance of high-quality leadership stressing that school leadership remains a key priority for Scottish education policy making (Scottish Executive, 2004; Scottish Executive, 2005). The Canadian Council of Atlantic Ministers of Education and Training (CAMET/CAMEF), a joint agency to foster exchange and collaboration in education in the provinces of New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island, has equally recognised the value of strong pedagogical leadership for effective teaching and learning as part of its 2008 Strategic Directions, making leadership standards and school leader development and appraisal an important objective (Fournier and Mildon, forthcoming). In the United States, a recent survey among policy makers, school and local administrators identified the leadership of school principals as the second most important area of national education reform after teacher quality (The Wallace Foundation, 2012).

Within that context, various countries have developed and implemented initiatives to strengthen school leadership by improving their approach to the recruitment and selection of school leaders and by providing support and professional development opportunities as part of their school improvement efforts (Pont et al., 2008a).

- In Chile, Congress passed the Quality and Equality of Education Law (Law No. 20.501) in January 2011 which introduced a wide range of policies to improve the quality of education. While the Quality and Equality of Education Law introduced policies related to teachers, it also recognised the importance of school leadership for effective schools through measures that aim to strengthen the professional status of principals in Chile. These include, among others, policies related to the selection and recruitment of principals (*Alta Dirección Pública*), the remuneration of principals (higher salaries according to school size and the number of underprivileged students enrolled), and greater autonomy for principals to organise leadership teams and to replace underperforming teachers. In addition, in 2011 Chile implemented a School Principals of Excellence training programme (*Programa Formación de Directores de Excelencia*).[14] This programme aims to support principals in their work and to develop skills for better school leadership among current and aspiring school principals through the provision of grants and scholarships to participate in high-quality, flexible and pedagogically-centred professional development programmes (e.g. through Master's programmes, Diploma programmes, and internships). Between 2011 and 2012, the Chilean government granted over 1 600 scholarships to fund the needs of teaching professionals keen to develop their school leadership skills. For 2013, 1 000 scholarships have been approved (Chilean Ministry of Education, forthcoming).
- New Zealand has invested considerably in developing school leadership competencies across its education system.[15] New Zealand's school leadership improvement efforts include a research-based model of effective pedagogical leadership, the Kiwi Leadership for Principals framework; the Educational Leadership Practices survey, a formative tool to help school principals analyse

their leadership in schools; and a Professional Leadership Plan offering professional development opportunities for school principals at different stages of their career (Nusche et al., 2012).

- Sweden introduced a National School Leadership Training Programme for beginning school leaders, deputy principals and principals of pre-schools in 2009.[16] Since 2010, beginning principals are required to enrol in this programme within the first year in office. The three-year programme, which is managed by the Swedish National Agency for Education and offered at universities, involves 32 days of residential training, work at the principal's own school and further study. The programme focuses on three areas of school leadership: 1) legislation and the role of exercising the functions of an authority; 2) management by goals and objectives; 3) school leadership. Universities receive funds to cover the costs for tuition, local education authorities to cover costs for seminars, travel and literature. Local education authorities are responsible for school leaders' salaries during the time of the training. Completion of the School Leadership Training programme enables principals to participate in the Principal Improvement Programme, which provides an opportunity to further develop pedagogical leadership competencies.

Initiatives such as these raise the question about the ways in which school leader appraisal can strengthen efforts to improve and develop school leadership. As Darling-Hammond and Rothman (2011) argued, policies for the development and support of both teachers and school leaders require a balanced approach including the recruitment of qualified individuals, their preparation, induction, professional development, appraisal, career development and retention over time. Some systems, such as Ontario, Canada, and Victoria, Australia, have firmly embedded school leader appraisal within a comprehensive school leadership strategy for school improvement (see Box 7.3).

Box 7.3 Embedding school leader appraisal within a comprehensive leadership framework and strategy: The cases of Victoria, Australia, and Ontario, Canada

Victoria, Australia

In 2003, Victoria introduced its state-wide reform agenda Blueprint for Government Schools. Based on the belief in the difference that school leaders can make for improving student outcomes and the role of school leaders as key drivers for making the wider reform agenda work, building leadership capacity represented an essential element of this reform programme. A Victorian leadership development strategy, Learning to Lead Effective Schools, was implemented in 2006. It included, among others, a Developmental Learning Framework for School Leaders, initiatives for an improved selection process for school principals, mentoring and coaching programmes, an accelerated development programme for high potential leaders, and a school principal performance management system. This Principal Class Performance and Development Process provides a framework that all school principals and assistant principals use in planning, reviewing and receiving feedback on school performance and leadership effectiveness. The process aims to provide school principals and assistant principals with objective and constructive feedback and learning opportunities. It is designed to support school leaders to develop the knowledge and skills necessary to lead sustainable and system-wide school improvement. A significant feature of the Victorian approach to school and system improvement is the high degree of alignment of all its strategies (e.g. leadership development programmes, selection processes and performance appraisal).

Box 7.3 Embedding school leader appraisal within a comprehensive leadership framework and strategy: The cases of Victoria, Australia, and Ontario, Canada (*continued*)

Since 2011, Victoria has been developing a new reform programme and set of strategic directions for education. These include 1) Towards Victoria as a Learning Community; 2) Refocusing Vocational training in Victoria; 3) New Directions for School Leadership and the Teaching Profession; 4) the Victorian Government's Vision for Languages Education. New Directions for School Leadership and the Teaching Profession constitutes a key element of the state government's goal to offer students a world-class education. A discussion paper released for consultation in June 2012 highlights the role of high-quality teaching for education. It identified support for school leadership at the school and system level as one of three policy levers to improve teaching in Victoria's schools. It puts forth three potential areas to further strengthen the leadership capacity of Victoria's schools: 1) the effective recruitment and preparation of school leaders; 2) support and appraisal; 3) system-level policies. Effective school leadership, then, remains at the heart of Victoria's efforts to improve teaching and learning for all students.

Sources: Matthews et al. (2008); State of Victoria, Department of Education and Early Childhood Development (2012), *www.education.vic.gov.au*.

Ontario, Canada

The province of Ontario, Canada, has identified successful school and system leadership as a core element of its efforts to achieve the province's three core educational goals: (1) high levels of student achievement; (2) reduced gaps in student achievement; and (3) increased public confidence in publicly funded education. To this end, Ontario has developed and implemented a comprehensive school and system leadership strategy, the Ontario Leadership Strategy (OLS), to support student achievement and well-being by attracting and developing skilled and passionate school and system leaders. As part of this strategy, several tools and support mechanisms (e.g. The Ontario Leadership Framework 2012, and Core Leadership Capacities) have been developed to streamline and focus efforts to support school principals and vice-principals, to refine leadership skills and to put advanced leadership concepts and practices to work on a daily basis to meet educational targets and achieve concrete results.

A province-wide Principal/Vice-Principal Performance Appraisal (PPA) system focussed on goals that promote student achievement and well-being constitutes a key component of the OLS. It is designed to support the strategy's two overarching goals: (1) to attract competent people to school leadership roles; and (2) to develop the best possible instructional leaders. With these goals in mind, Ontario's PPA has been designed to ensure that school leaders are well-supported in their development through targeted, system-wide strategies that provide formal and informal opportunities for feedback and ongoing professional development. During the performance appraisal process, school principals and vice-principals develop goals to improve student achievement and well-being based on the board's improvement plan, the school improvement plan, the school community and local context, ministry priorities, and personal growth and development goals. Ontario has also been developing and piloting an appraisal system for supervisory officers.

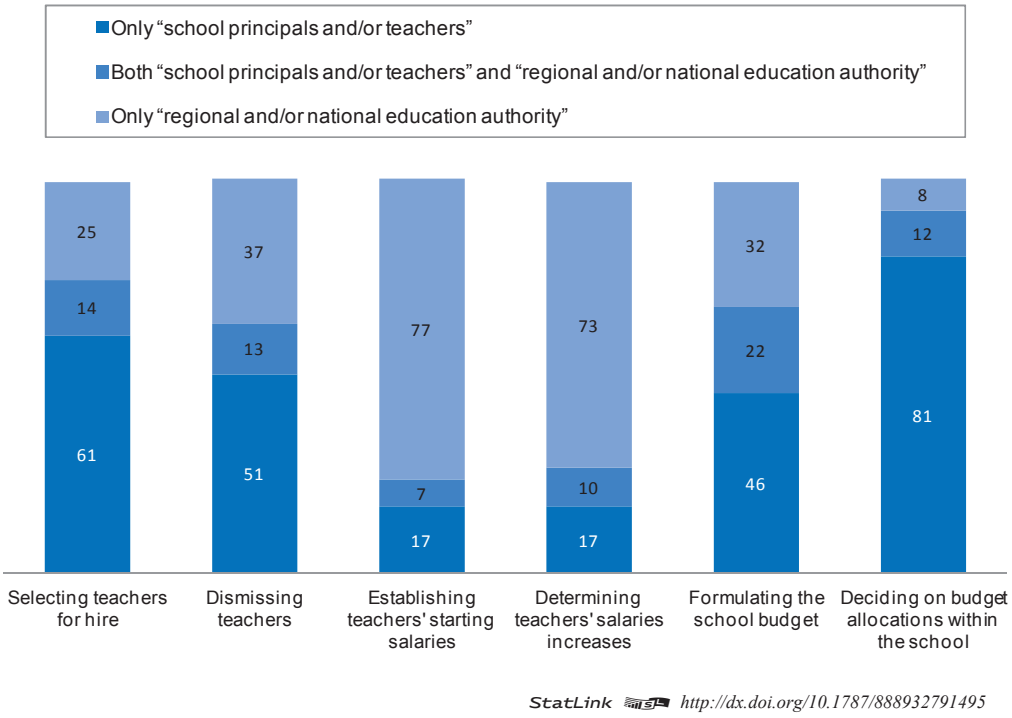
Sources: Pervin and Campbell (2011); Ontario Leadership Strategy, *www.ontario.ca/eduleadership;* Ontario Institute for Education Leadership website, *www.education-leadership-ontario.ca/content/home*.

Leadership roles are becoming more complex

In their efforts to strengthen education systems and improve student learning, various countries have decentralised decision making in education in different ways (also see Chapter 2). Depending on the context, intermediate levels of governance (e.g. regions, provinces and local education authorities) and/or schools and school boards have taken over responsibilities related to teaching and learning (e.g. curriculum, assessment and instruction) and/or the governance and management of schools (e.g. human resources and school funding). In various countries, school principals and, at times, department heads and teachers, hold great decision making powers in some areas of resource allocation and curricula and assessment (OECD, 2010, see Figures 7.2 and 7.3). Most countries, however, combine different elements of centralisation and decentralisation with substantial interplay between the different levels of the education system (Pont et al., 2008a).

Figure 7.2 School autonomy over resource allocation (2009)

OECD average for students in schools whose school principals reported that the following actors have a considerable responsibility for the following task (%)

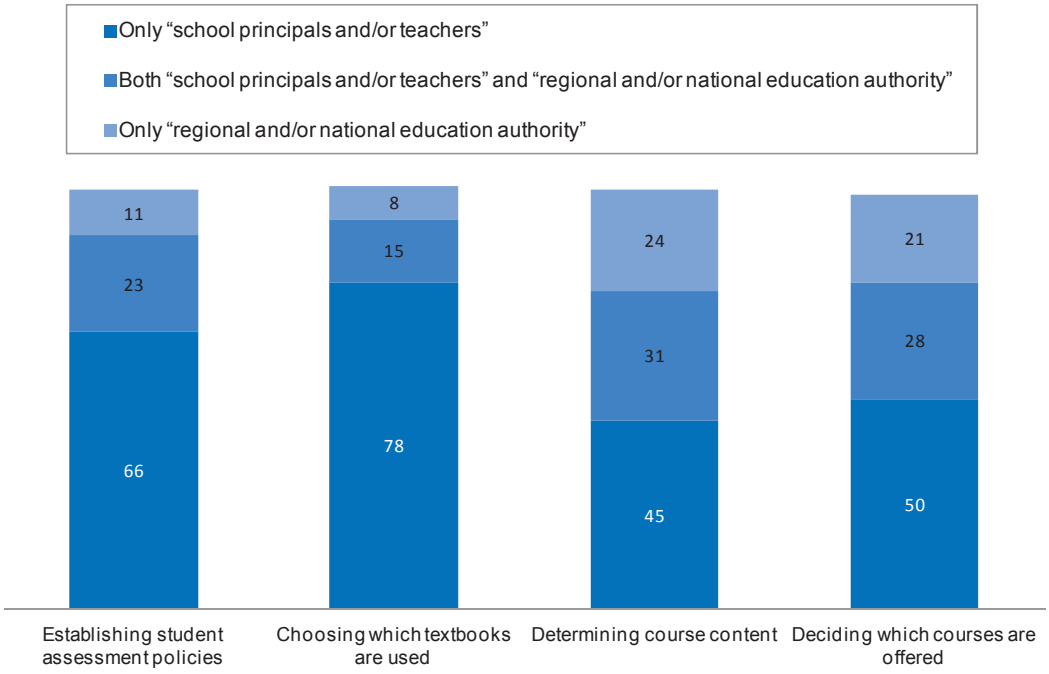


StatLink http://dx.doi.org/10.1787/888932791495

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

Figure 7.3 School autonomy over curricula and assessment (2009)

OECD average for students in schools whose school principals reported that the following actors have a considerable responsibility for the following tasks (%)



StatLink http://dx.doi.org/10.1787/888932791514

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

Even though greater school autonomy does not necessarily lead to a change in school principals' leadership styles, as the OECD Teaching and Learning International Survey (TALIS) school principal questionnaire administered in 23 countries indicated (OECD, 2009), more local powers in educational decision making have changed leadership roles and responsibilities and heightened the need for effective leadership in schools. Eurydice, for example, pointed out that within a mixed landscape of school autonomy across Europe, in countries that have delegated responsibilities for human resource management (e.g. teacher recruitment, defining teacher responsibilities, and granting additional salary payments) to the school level, school principals are most often responsible for these new tasks. In a majority of the countries covered by Eurydice, school heads, either alone or in collaboration with their school board, were also responsible for decisions related to the curriculum content of optional subjects and the grouping of students for compulsory learning activities (EACEA, 2012). Greater decentralisation and school autonomy, then, often coupled with the introduction of stronger accountability frameworks and an increasing workload, demand new skills and competencies from school leaders (Leithwood, 2001; Normore, 2003; Pont et al., 2008a; Louis and Robinson, 2012). This includes the demand for new skills and competencies stemming from the implementation of comprehensive evaluation and assessment frameworks and school leaders' growing responsibilities for teacher appraisal and the self-evaluation of their school (see Chapters 3, 5 and 6).

Besides changes in educational governance, wider changes in society, such as technological changes, globalisation, migration and increasing diversity also influence education and the kind of leadership that schools require. In various countries, policies that seek to respond to these changes aim to increase the focus of schools on teaching and learning. These social changes and related policies as well as new understandings about student learning and instruction demand more from school leaders than the administrative implementation of rules and regulations (Pont et al., 2008a; OECD, 2009; Parylo et al., 2012b). In Hungary, for example, schools and school leaders face the challenge of integrating competence-based education, a growing focus on the learning of foreign languages and the use of new technologies into their classrooms (Halász, 2009). In the Netherlands, increasing individualisation together with increased diversity in society place new demands on schools and school leaders to play an active role in promoting citizenship and social integration (Bal and de Jong, 2007).

The effects of greater decentralisation and school autonomy as well as wider changes in society and education on school leaders' roles and required competencies, in turn, demand stronger support systems to prepare school leaders for their new tasks before holding them accountable for their greater responsibilities. This has important implications for the ways in which appraisal is designed to ensure support as well as accountability.

There are concerns about existing approaches to school leader appraisal.

Education researchers and policy makers have become increasingly aware of the shortcomings of current approaches to the appraisal of school leaders. While many of the countries that took part in the OECD project on Improving School Leadership had some form of school leader appraisal in place, many reported concerns about the scarcity of sound tools and mechanisms to best monitor and assess school leaders' performance (Pont et al., 2008a). In the United States, education researchers since the early 1990s have stressed the shortcomings of school principal appraisal systems (Ginsberg and Berry, 1990; Ginsberg and Thompson, 1992; Stufflebeam and Nevo, 1993). Recent reviews of school district approaches to school principal appraisal in the United States have echoed these earlier concerns about appraisal procedures in terms of their usefulness, accuracy, fairness and focus on leadership practices linked to better teaching and learning (Portin et al., 2006; Goldring et al., 2008; Davis et al., 2011; Parylo et al., 2012b). Even though school leader appraisal involves inherent difficulties owing to the complex and highly context-dependent nature of school leadership (Ginsberg and Thompson, 1992), a growing body of research in the United States has recently explored ways to create effective school leader appraisal systems (see, for example, Murphy et al., 2006; Porter et al., 2006; Goldring et al., 2007; The Wallace Foundation, 2009; Reeves, 2009; New Leaders for New Schools, 2010; Clifford and Ross, 2011; Sanders and Kearney, 2011).

The school leadership profession is faced with numerous challenges.

The current state of the school leadership profession influences considerations about school leader appraisal. The OECD project on Improving School Leadership (Pont et al., 2008a) shed light on the current challenges that school leaders as well as education systems face in various contexts. As Pont et al. pointed out, in various contexts the school leadership profession confronts an imminent wave of retirements of many school principals over the next five to ten years, difficulties to attract suitable candidates, a negative image resulting from difficult working conditions, an increasing workload and high stress levels, a lack of preparation and training, inadequate salaries and rewards and the need to adapt to changing models of school leadership. Mulford (2003) equally

stressed that school leaders in different contexts face challenges concerning their overall well-being and work-life balance, as well as feelings of professional isolation, often exacerbated by a lack of support provisions. To give an example, in the Flemish Community of Belgium, the School Advisory Services highlighted the challenge of recruiting school principals in light of a decreasing number of candidates and a growing workload for school principals, which, in turn, makes school leadership positions even less attractive (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

These challenges need to be taken into account when designing and implementing school leader appraisal frameworks. Depending on the design and implementation of appraisal, appraisal may, on the one hand, constitute one policy lever for addressing the current and future challenges as part of wider development and support systems for school leaders. On the other hand, appraisal bears the risk of exacerbating the challenges for school leaders further. Normore (2004) and Pashiardis and Brauckman (2008), for example, pointed towards these risk of accountability measures, which may lead to role overload and role ambiguity, higher levels of stress and professional burn out among school leaders.

Following the conceptual framework described above, the following sections describe the features of school leader appraisal, starting with the governance of appraisal schemes. Table 7.A.1 provides a concise overview of school leader appraisal systems and their characteristics in countries participating in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

Governance

This section analyses requirements for the appraisal of school leaders, the distribution of responsibilities for the design of appraisal frameworks and the objectives of appraisal. A taxonomy of country approaches to the appraisal of school principals in public schools developed on the basis of information gathered from countries participating in the OECD Review provides an overview of the governance of countries' appraisal systems.

Setting requirements and distributing responsibilities for the design of school leader appraisal

Approaches to appraisal are most often embedded within the overall structures of governance in education and often depend on the education authorities responsible for school leaders as their employer (see Tables 7.A.1 and 7.A.2). While school leaders in various countries are not appraised (e.g. Austria, Iceland, Ireland, Israel [ISCED level 3], Italy and Luxembourg), school leaders in a number of countries are appraised in various ways.[17] Education authorities at different levels of government play a major role in setting the requirements and in defining school leader appraisal procedures (e.g. for the definition of appraisal aspects and criteria, the development of professional school leadership standards and the design of related appraisal tools and instruments). Based on the information submitted by countries participating in this OECD Review, countries in which school principals in public schools are appraised can be broadly grouped according to four approaches depending on the requirements for appraisal and the locus of agency for its design and implementation. The Korean Evaluation for School Management system constitutes a specific case with regards to requirements for appraisal. In the case of the Korean Evaluation for School Management system, it is at the discretion of metropolitan and provincial offices of education to implement the central appraisal

framework or not. Table 7.1 illustrates the different approaches for the appraisal of school principals in public schools in the form of a taxonomy.

Table 7.1 A taxonomy of country approaches to the appraisal of school principals in public schools

		Agency for the design and implementation of appraisal			
Requirements for appraisal		Central/state framework for appraisal procedures, implementation through central/state, regional or intermediate authorities	Central/state framework for appraisal procedures, implementation through local authorities and/or school organising bodies	Appraisal procedures determined by local authorities and/or school organising bodies	No appraisal
	Mandatory periodic appraisal	Various states and territories in Australia, Belgium (French Community), France (ISCED levels 1, 2 and 3), Israel (ISCED level 1 and partially ISCED level 2), Korea (School Principal Appraisal for Professional Development), Mexico (centrally managed schools at ISCED level 3), Portugal, various autonomous communities in Spain	Various states and territories in Australia, Belgium (Flemish Community), various provinces and territories in Canada (e.g. Ontario), Chile (Performance Appraisal), France (ISCED level 1), New Zealand, Slovenia, United Kingdom (Northern Ireland)		
	Mandatory appraisal linked exclusively to employment-related decisions (contract renewal)	Poland	Poland	Czech Republic, Slovak Republic	
	Appraisal at the discretion of regional education authorities	Korea (Evaluation for School Management)			
	Appraisal at the discretion of local authorities and/or school organising bodies			Various provinces and territories in Canada (e.g. Alberta, Newfoundland and Labrador, New Brunswick, Nova Scotia, and Prince Edward Island), Czech Republic, Denmark, Estonia, Finland, Hungary, Netherlands, Norway, Slovak Republic, Sweden	
	No appraisal				Austria, Iceland, Ireland, Israel (ISCED level 3), Italy, Luxembourg

Note: This taxonomy only includes countries which submitted information based on a questionnaire developed by the OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes. Countries with various requirements for the appraisal of school principals and countries with multiple jurisdictions have been included in all respective cells to reflect differences across sub-national jurisdictions.

In a first group of countries, there is no central/state requirement to appraise school principals and central/state education authorities have not established policy frameworks that set out procedures for appraisal. Instead, school principal appraisal typically takes place entirely at a local level. School principals are appraised at the discretion of local education authorities and/or school organising bodies, which also determine and implement appraisal procedures. This group includes some provinces and territories in Canada, the Czech Republic, Denmark, Estonia, Finland, Hungary,[18] the Netherlands,

Norway, the Slovak Republic and Sweden (see Box 7.4 for examples of appraisal procedures as defined and implemented at a local level).

In a second group of countries, appraisal is required as an employment-related process in the case of re-appointing school principals into their position. This includes the Czech Republic, Poland and the Slovak Republic. The locus of agency for designing and implementing procedures for this kind of appraisal differs between countries. School principals in Poland are appraised by local and regional education authorities according to a central framework. In the Czech Republic and the Slovak Republic, school organising bodies hold responsibility for determining and implementing appraisal procedures, as is the case for the informal appraisal process at a local level in both countries.

In a third group of countries, school principals are appraised on a mandatory periodic basis according to procedures defined through a central/state framework for local implementation with scope for adaptation and decision making at the local level (e.g. by local education authorities, school organising bodies and school boards). Depending on the context, requirements for appraisal may additionally be linked to employment-related requirements. Eight systems reported such appraisal approaches: various states and territories in Australia, Belgium (Flemish Community), various provinces and territories in Canada, Chile (Performance Appraisal), France (ISCED level 1), New Zealand, Slovenia and Northern Ireland in the United Kingdom.

In a fourth group of countries, school principals are appraised on a mandatory periodic basis according to procedures defined through a central/state framework that are implemented through central/state, regional or intermediate authorities. Depending on the context, requirements for appraisal may, as in the third group, additionally be linked to employment-related requirements. In three systems (various states and territories in Australia, Belgium [French Community] and Mexico [centrally managed schools at ISCED level 3]) appraisal is implemented by central/state education authorities. In five countries (France [ISCED levels 1, 2 and 3], Israel [ISCED level 1 and partially ISCED level 2], Korea [School Principal Appraisal for Professional Development], Portugal and various autonomous communities in Spain) intermediate or regional education authorities are responsible for implementing the central/state appraisal framework.

However, the extent to which central/state frameworks prescribe procedures for the appraisal of school principals may differ widely (e.g. through setting professional school leadership standards). The Northern Territory and Victoria, Australia; Ontario, Canada; and Chile represent examples of systems with very detailed requirements and guidelines on how appraisal is to be carried out at a local level.

- In the Northern Territory and Victoria, Australia, state and territory authorities provide a very detailed framework for the implementation of school principal appraisal at a local level. In the Northern Territory, the Department of Education and Training has developed a School Accountability and Performance Improvement Framework that sets out detailed requirements for the appraisal of school principals. A range of guides, templates and tools seeks to facilitate implementation by local evaluators, the Directors School Performance, that ultimately determine the specific approach to appraisal. The Victorian Office for Government School Education of the Department of Education and Early Childhood Development has also developed detailed guidelines for the Principal Class Performance and Development process that lay out the requirements for school principal appraisal and guide regional directors and local designated officers in the appraisal process.

- In Ontario, Canada, school boards made up of supervisory officers and directors of education implement the province's performance appraisal system for school principals that has been designed and conceptualised by the provincial education authorities. The ministry provides school boards with a mandatory summative report form and sample working templates for the Performance Plan, the Annual Growth Plan and the Improvement Plan that may be modified to suit local needs within the legal requirements.
- In Chile, central requirements and guidelines for the Performance Appraisal process are implemented by the municipal school organising bodies. The Chilean central authorities provide detailed guidelines and instructions on a dedicated website as well as an interactive online platform for the implementation of appraisal.

Depending on the context, among countries that require the mandatory periodic or employment-related appraisal of school principals, these requirements may differ between public and private schools. In various countries, requirements only apply to public schools (e.g. various states and territories in Australia, Chile [Performance Appraisal], Czech Republic, France (ISCED levels 1, 2 and 3), Mexico [centrally managed schools at ISCED level 3], Poland, Portugal, and various autonomous communities in Spain). In France, for example, school principals of private schools are appraised by their employer, typically in relation to the description of their general and professional duties. However, school principals in French private primary schools are also subject to an appraisal by local inspection bodies (*Inspecteur de l'éducation nationale*) that focuses on school leaders' compliance with national education goals and programmes (Dos Santos and Rakocevic, 2012). In various other countries, requirements apply to both public and government-dependent private schools (e.g. Belgium [Flemish and French Communities] and Slovenia) or all schools (e.g. Israel [ISCED level 1 and partially ISCED level 2],[19] Korea [School Principal Appraisal for Professional Development], New Zealand, the Slovak Republic, and Northern Ireland in the United Kingdom).

Requirements may also vary between school levels and school types reflecting governance structures in education (e.g. France, Israel and Mexico). In France, appraisal requirements, responsibilities and procedures differ for school principals in primary and secondary schools. This also reflects respective differences in responsibilities. The Israeli appraisal requirements under the New Horizon Reform only apply to schools at the primary level and partially to schools at the upper secondary level. The Mexican appraisal requirements described in this chapter only apply to schools at the upper secondary level that are managed by central education authorities.

Deputy school principals are appraised according to the same requirements in various states and territories in Australia, Belgium (French Community), in various provinces and territories in Canada, Chile (Performance Appraisal), France (ISCED levels 1, 2 and 3), Israel (ISCED level 1 and partially ISCED level 2), Korea (School Principal Appraisal for Professional Development), Slovak Republic, and Northern Ireland in the United Kingdom.

Box 7.4 School leader appraisal procedures at a local level

Atlantic provinces, Canada

Based on a strategy developed by the Council of Atlantic Ministers of Education and Training (CAMET/CAMEF), the Atlantic provinces of Canada, New Brunswick, Newfoundland and Labrador, Nova Scotia and Prince Edward Island, have recently established a set of professional standards for school leaders to guide performance appraisal by school boards and boards of education. In the Atlantic provinces, school principals are typically appraised by superintendents or directors of education according to procedures set by their responsible school board or board of education. Superintendents and directors of education usually receive training related to appraisal. Some boards require an annual appraisal of new school principals based on three meetings per year. Other boards have implemented a three-year appraisal cycle that involves a process of establishing goals, related indicators of achievement and an overall timeline. School principals and evaluators meet periodically throughout the school year to review a school leader's progress and revise objectives as necessary. At the end of the school year, school principals and evaluators meet to discuss results and the follow-up for the subsequent year. At the end of the 3-year appraisal cycle, evaluators provide a statement of a school principal's performance that is reviewed and signed by the school principal and added to the school principal's human resource files. Regarding the appraisal procedures, tools and instruments, several of the school boards and boards of education of the Atlantic provinces require school leaders to reflect on their own performance. In the Eastern District of Prince Edward Island school principals are, for example, asked to provide a self-appraisal according to specified leadership tasks and responsibilities (e.g. curriculum and instruction, staff supervision and development, parent and community involvement, operational management and impact of work on school improvement plan). School boards or boards of education in the Atlantic provinces often also require the collection of stakeholder views as part of the appraisal. In the Nova Central District in Newfoundland and Labrador, stakeholder views feed into the appraisal of school leaders based on the rationale that school leaders need support and feedback from teachers as well as colleagues, students, parents and support staff. In the Eastern District of Prince Edward Island, teachers and parents are asked to complete a questionnaire on their perspective on the quality of school leadership.

Denmark

In Denmark, school leaders of public schools at primary and lower secondary levels are typically appraised by their responsible municipality. Practices, however, vary between schools and municipalities. Some municipalities require an annual appraisal of school leaders. Increasingly, school principals are appraised through a management by means of objectives approach. Even though primarily a management tool, this approach equally serves as a means to hold schools and school leaders accountable for their performance. The management by objectives approach typically requires school principals to establish a set of objectives for the school as a whole as well as their own performance (e.g. through a results contract or a school principal agreement), usually for a one to two year period, that are, then, continuously monitored. No comprehensive overview of the instruments used for the continuous assessment of goals and objectives is available. School principals in public schools at upper secondary level are also typically held accountable through annual results contracts.

Slovak Republic

School organising bodies (e.g. municipalities, regional self-administration authorities, religious bodies) are legally required to appraise school headmasters in relation to their contract renewal and hold responsibility for determining appraisal procedures. School headmasters are typically appraised on their management responsibilities and competencies through an individual interview. Appraisal often includes several of the following aspects and criteria: knowledge of and compliance with legislative requirements, financial management, school development, student outcomes, the level of teamwork in a school and a school's collaboration with its external partners. Pedagogical leadership and the quality of education in a school are typically not part of the appraisal. The sources of information that school organising bodies rely on often include: statistical information, school reports, budget figures and the results of external school evaluations by the school inspectorate. Appraisal results are sometimes used to reward school headmasters through a letter of recognition or a celebration as part of the teachers' day. Results can also be used as a factor for determining the variable part of a school headmaster's salary, but this influence is minimal and only ranges between 0-30%. Appraisal results are not typically used to inform school headmasters' professional development, which often focuses on administrative and management tasks (e.g. changes in legislation and health and safety at work).

Sources: Fournier and Mildon (forthcoming); Danish Ministry of Education and Rambøll (2012); Hajdúková et al. (forthcoming).

There are trade-offs between consistency through central/state frameworks and local diversity

Systems that leave the appraisal of school leaders entirely to the discretion of local education authorities and systems that require the appraisal of school leaders, but that lack central/state frameworks for procedures, allow for the greatest possible degree of local adaptation. Considering the growing realisation that successful school leadership is always context-dependent, there are strong arguments for a fair degree of local agency in the design of school leader appraisal. The local design of appraisal procedures also constitutes an opportunity to create trust, commitment and ownership in the appraisal process and provides opportunities for the development of innovative approaches at a local level. Depending on the system and the availability of opportunities for co-operation, these innovations and good practices could be shared between local actors.

However, in systems which allow local education authorities complete autonomy to decide whether to appraise their school leaders and how to design appraisal processes there may be few expectations that school leaders are accountable for the quality of teaching and learning and the student outcomes in their school. In systems that allow local education authorities complete autonomy about appraisal requirements and procedures or that require appraisal, but lack central/state frameworks for procedures, appraisal processes may also differ considerably between local authorities. Thomas et al.'s (2000) study, for example, showed that the appraisal approaches used in Alberta, Canada, were often unique to each school district. In systems with a high degree of local responsibility for the design of appraisal not all school leaders might, therefore, experience an appraisal that they perceive as useful and meaningful (Kimball et al., 2009). In Sweden, for example, the Association of School Principals and Directors of Education (*Sveriges Skolledarförbund*) raised concerns about municipalities not devoting enough time and resources to the appraisal of school leaders (Nusche et al., 2011). The OECD Country Review of the Czech Republic similarly identified challenges for ensuring an effective appraisal process of school principals through school organising bodies. The study found that, in many cases, the appraisal of school principals fails to go beyond a compliance-driven process that focuses on the financial aspects of budget management (Santiago et al., 2012a). Local differences may also lead to a fairness issue if not all school leaders participate in an equally rigorous appraisal process, especially where appraisal may lead to summative consequences and feeds into employment- and career-related decisions. Furthermore, locally developed appraisal systems risk a lack of alignment with research on effective school leadership and school leader appraisal (Davis et al., 2011). Scope for local agency in the design of appraisal may, then, cause concerns of accuracy, utility, validity, reliability and fairness (see Box 7.5 for an explanation of the required properties of school leader appraisal).

One key challenge, therefore, concerns the need to ensure that all school leaders benefit from effective appraisal and feedback arrangements that reflect local needs at the same time. The distribution of responsibilities needs to strike the right balance between centrally/state-directed policies, on the one hand, and scope to meet local needs, on the other.

Box 7.5 Properties of school leader appraisal

The research literature has identified a number of different, often interrelated, properties that appraisal needs to fulfil.

Ginsberg and Berry (1990) highlighted the role of reliability, validity and utility concerns. As they stressed, the level of reliability and validity that is required depends on the uses of results and high levels of reliability and validity are particularly important for summative decisions.

- Reliability refers to the consistency of measurements across evaluators and observations. It describes the extent to which an assessment or evaluation yields the information it is designed to capture and return the same results for assessments or evaluations in similar conditions across schools and over time.
- Validity describes the extent to which an assessment or evaluation is designed so that it supports its intended purposes and the extent to which results allow for meaningful inferences to support decisions and uses. It describes the comprehensiveness in assessing performance according to set aspects and criteria. Factors such as the clarity of aspects and criteria, the sources of information and the capacity of evaluators all influence the level of validity.
- Utility, related to reliability and validity, describes the consistency and accuracy with which appraisal measures different degrees of performance. It also refers to the balance between benefits (e.g. improving decision making processes and performance) and costs (e.g. logistical and financial).

Stufflebeam and Nevo (1993) described a different set of features that appraisal needs to fulfil. According to Stufflebeam and Nevo, and based on the Joint Committee's Personnel Evaluation Standards in the United States, appraisal should fulfil four basic requirements:

- Propriety, i.e. due attention to the rights and welfare of the persons affected by an appraisal
- Utility, i.e. the balance between costs and benefits and the extent to which appraisal fulfils its purposes
- Feasibility, i.e. the extent to which appraisal is easy to implement and administer
- Accuracy, i.e. the extent to which conclusions are drawn logically from the data.

Reeves (2009), furthermore, stressed the need for objectivity and fairness, i.e. the need for appraisal to rely on description and not conjecture. However, as Pashiardis and Brauckman (2008) pointed out, even the most objective data can be subjectively interpreted. As they, therefore, argued, evaluation, assessment and appraisal can only try to reduce inevitable subjectivity.

Objectives

Similar to the wider evaluation and assessment framework, the overarching objective of school leader appraisal is ideally aligned with the overall goals of schooling in a particular context (also see Chapter 3). Within the respective general objectives, appraisal is directed towards improved teaching and learning for all students (Reeves, 2009; Leon et al., 2011). The overall goals with which school leader appraisal is aligned can be achieved through two major functions (Goldring et al., 2007). First, appraisal can seek to improve school leaders' practices by identifying strengths and weaknesses. *Developmental appraisal* can provide feedback for the improvement of school leaders' practices and behaviours. It can provide a platform for school leaders to learn about, reflect on and improve their practices and behaviours and to inform professional development activities. Second, school leader appraisal can aim to hold school leaders accountable for their performance and the quality of education in their school. *Appraisal for accountability* can inform employment- and career-related decisions in order to set incentives for school leaders to perform at their best. It may entail performance-based career advancement and financial rewards (e.g. bonus pay) for outstanding school leaders as well as the possibility of sanctions for underperformance. School leader appraisal can also constitute an opportunity for identifying effective leaders to celebrate and reward their achievements (see Table 7.A.1 for a general overview).

Some countries primarily aim to hold school leaders accountable through their appraisal system. In the Czech Republic, Poland and the Slovak Republic requirements to appraise school leaders are linked exclusively to employment-related decisions and for informing contract renewal processes. In France (ISCED levels 1, 2 and 3), Portugal and various autonomous communities in Spain, school leader appraisal also serves a predominantly summative function. In all of these three countries, appraisal is designed to mainly hold school leaders accountable for their performance and to inform decisions about their career advancement, rewards and/or consequences for underperformance.

A number of countries are moving towards appraisal for improvement through a combination of developmental and accountability functions or a sole focus on professional development. Systems that combine accountability and developmental functions and that use appraisal results for different summative and formative decisions include various states and territories in Australia (e.g. Northern Territory and Victoria), Belgium (Flemish Community), various provinces and territories in Canada (e.g. Ontario), Israel (ISCED level 1 and partially ISCED level 2), Mexico (centrally managed schools at ISCED level 3), New Zealand and Northern Ireland in the United Kingdom (see Box 7.6 for examples for the ways in which systems combine developmental and accountability functions within one framework). In Belgium (French Community) appraisal serves a solely formative purpose and appraisal does not result in any summative consequences for school principals and deputy principals. A few countries have appraisal schemes in place that are designed to reward successful leadership. Slovenia constitutes one of the few examples of such appraisal systems for rewards purposes. In Slovenia, the appraisal of school leaders serves as an opportunity to identify and reward successful school leaders with a one-off financial bonus.

Box 7.6 Approaches to combining developmental and accountability functions within one appraisal framework

Some systems combine developmental and accountability functions within one appraisal framework through the overall timeframe for appraisal. As part of this approach, school leaders are typically appraised over several years through a formative appraisal cycle. At the end of the formative appraisal cycle, appraisal typically informs summative decisions, such as a school principal's contract renewal process.

- In the Northern Territory, Australia, school principals are generally employed on a fixed-term contract of four years. School principals are appraised in a formative process over a period of 18 months that includes coaching conversations between school principals and evaluators after 6 and 12 months. At the end of the formative appraisal period, results of appraisal inform decisions about a school principal's future career, including the contract renewal process. A similar procedure is in place in the Australian state of Victoria.
- In the Western District in the province of Prince Edward Island, Canada, appraisal equally includes a formative appraisal through an annual goal setting process that provides an opportunity for professional growth. At least every four years, the same process has a summative dimension that informs the contract renewal process.

Other systems combine developmental and accountability functions within one appraisal framework through the possibility of initiating a separate process or a further appraisal cycle that may lead to summative consequences in the case of a school leader's continued underperformance.

- In Ontario, Canada, the Principal and Vice-Principal Performance Appraisal system aims to provide support to improve school leaders' practices and behaviours. However, following a second or third unsatisfactory appraisal, a school leader may be dismissed or re-assigned to a different position.
- The appraisal procedures that have come into effect in England in the United Kingdom in September 2012 aim to provide the basis for a supportive and developmental process to ensure that headteachers have the skills and support they need to carry out their role effectively. The process may, however, also give recommendations on pay and, in the case of concerns about a headteacher's performance, result in the initiation of a separate capability procedure to decide about the dismissal of a headteacher.
- Northern Ireland in the United Kingdom follows a similar approach. In case of an unsatisfactory result in the Performance Review and Staff Development Scheme, school principals and vice principals receive informal support and development opportunities. This may be followed by a formal stage which includes the issue of a formal written notice and a targeted support programme. If a principal or vice principal fail to reach a satisfactory level of performance after the informal and formal stages of support, he or she may be ultimately dismissed.

Sources: Department of Education and Children's Services, Northern Territory Government, Accountability and Performance Improvement Framework (APIF), *www.det.nt.gov.au/about-us/APIF*;
Victoria Department of Education and Early Childhood Development Principal Class Performance and Development Process, *www.education.vic.gov.au/school/principals/profdev/Pages/principalpd.aspx*;
Prince Edward Island Department of Education and Early Childhood Development, Western School Board of Prince Edward Island, School Principal Supervision & Evaluation Process,
www.edu.pe.ca/wsb/schoolboard/policiesproceeduresguidelines/communitiesforlearning/35-Supervision&Evaluation.pdf;
The Education (School Teachers' Appraisal) (England) Regulations 2012 No. 115,
www.legislation.gov.uk/uksi/2012/115/pdfs/uksi_20120115_en.pdf;
Ontario Ministry of Education website, Leadership Development, Principal Performance Appraisal,
www.edu.gov.on.ca/eng/policyfunding/leadership/appraise.html,
Department of Education, Northern Ireland (forthcoming).

Two countries, Chile and Korea, represent innovative cases for introducing two separate appraisal schemes for different purposes (see Box 7.7).

Box 7.7 The introduction of two separate appraisal schemes for different purposes: The cases of Chile and Korea

Chile

Chile has introduced two separate appraisal processes for school principals and teaching and technical-pedagogical leaders. One of the appraisal systems, the Performance Appraisal system (*Evaluación de Desempeño*), represents an annual performance management process that is primarily designed to foster reflection and inform professional development. The second system, Appraisal of Collective Performance (*Asignación de Desempeño Colectivo*), constitutes a voluntary performance management system that aims to foster distributed leadership and reward successful school leadership teams (see Box 7.2 for further details). Both processes, the Performance Appraisal system and the Appraisal of Collective Performance process, can also inform each other through the setting of the same institutional objective and targets, an essential element of both appraisal systems.

Sources: Chilean Ministry of Education (forthcoming), El Portal Gestión y Liderazgo Educativo, *www2.gestionyliderazgoeducativo.cl*, and Más Directivos, *www.masdirectivos.cl*.

Korea

Korea has developed and implemented two appraisal systems for school principals: (1) the summative Evaluation for School Management process and (2) the formative School Principal Appraisal for Professional Development system as part of the Teacher Appraisal for Professional Development scheme. Results of the Evaluation for School Management scheme are typically used to provide benefits for the top performers in appraisal through career opportunities, promotions and performance-based rewards, and to impose sanctions on those rated among the bottom 2-3%. In contrast, results of the School Principal Appraisal for Professional Development scheme are mostly used to further develop principal professional capability and as evidence material to request local education authorities to develop improvement policies. The Evaluation for School Management has been implemented by various metropolitan and provincial offices of education. As of February 2010, 6 out of the 16 metropolitan and provincial offices of education were conducting or had announced plans to conduct the school management appraisal process. The School Principal Appraisal for Professional Development process has been carried out by all metropolitan and provincial offices of education since 2010.

Source: Kim et al. (2010).

There are benefits to combining appraisal for professional development and accountability, but this approach involves difficulties

A study by Sun and Youngs (2009) on the ways in which school principal appraisal affects pedagogical leadership in 13 school districts in Michigan, United States, found that the school principals that took part in the study were more likely to engage in pedagogical leadership if appraisal aimed to inform professional development, to encourage school restructuring and to hold school principals accountable for student learning. Based on their findings, Sun and Youngs argued that a combination of developmental and accountability functions as the objectives of appraisal may help ensure that appraisal not only assesses pedagogical leadership in its own right, but with a view towards its impact on student learning. As they suggested, a combination of developmental and accountability purposes may also ensure that professional development activities are targeted at valued outcomes. Sun and Youngs concluded that, in order to effectively influence school leaders' practices and behaviours, appraisal needs

to go beyond employment-related decisions (e.g. contract renewal or salary allocations) and, instead, promote professional development, accountability and school restructuring. A similar study by Sun et al. (2012) on the appraisal of school principals in Michigan, United States, and Beijing, China, gave similar insights into the combination of developmental and accountability functions.

However, the combination of developmental and accountability functions involves challenges that may jeopardise the effective implementation of appraisal. The definition of appraisal purposes influences the design of appraisal procedures, the information the appraisal process seeks to gather for a specific target group, the different tools and methods used to do so, and the ways in which appraisal results are used.

- Pashiardis and Brauckman (2008) argued that formative appraisal is typically more oriented towards processes and the needs of the person appraised, while summative appraisal focuses on outcomes and the needs of the wider system. Summative appraisal equally mainly seeks to assess a leader's competences and achievements without a vision for future development. Formative appraisal, on the other hand, is rather oriented towards informing future action and growth (Condon and Clifford, 2010).
- The purposes also influence the relationship between appraisers and school leaders as well as the kind of information a school leader is likely to reveal. Both, developmental and accountability functions, may influence a school leader's willingness for co-operation in the appraisal process and the relationship and trust that can be established between appraiser and appraisee. Appraisal for professional development requires school leaders' openness towards their own weaknesses. Appraisal for accountability encourages school leaders to emphasise their strengths (The Wallace Foundation, 2009).
- Development and accountability demands raise different requirements for the involvement of school leaders in their own appraisal. Appraisal schemes that allow great scope for school leaders to influence appraisal procedures (e.g. the definition of aspects and criteria or the selection of sources of information), may lead to fairness concerns if appraisal involves summative consequences (Cullen, 1997).

Policy makers, therefore, face the challenge of finding an approach to appraisal that ensures that school leaders receive regular feedback and support, but that school leaders are also held accountable for the quality of education in their schools. Policy makers need to find ways to ensure that the combination of different appraisal purposes does not undermine the appraisal process. This may require policy trade-offs concerning the ways in which the appraisal system is designed.

The effective appraisal of school leaders requires clear and agreed objectives

Studies in different countries have highlighted the risk of different perceptions of appraisal objectives between evaluators and school leaders. Research examining and comparing evaluators' and school leaders' views of appraisal in Alberta, Canada (Thomas et al., 2000),[20] Israel (Gaziel, 2008) [21] and California, United States (David and Hensley, 1999)[22] revealed that school leaders and evaluators often do not necessarily share the same perceptions and expectations of the objectives of appraisal. The views of both school leaders and evaluators did also not necessarily match with the purposes intended by policy makers.

- Thomas et al.'s (2000) qualitative study of school principal appraisal in Alberta revealed substantial differences in the perceptions and attitudes towards the purposes of appraisal between superintendents and school principals. While most often designed by the individual school districts to promote professional growth, to provide information for administrative decisions and to clarify role expectations, the school principals involved in the study only shared the objective of promoting professional growth as one of their most important functions of appraisal. School principals and evaluators also only agreed on the general need of assessing school principals' effectiveness as one of their four most highly ranked purposes of appraisal.
- A qualitative study of primary school principals and school supervisors in Israel found that, among the study's participants, school principals expected appraisal to promote their professional development as the most important purpose of appraisal, while evaluators expected appraisal to fulfil a number of different purposes, including providing public accountability and informing decisions about school principals' promotions (Gaziel, 2008). Overall, Gaziel found that, among the study participants, evaluators perceived appraisal as more oriented towards accountability, school principals towards professional development.
- Davis and Hensley's study (1999) similarly found that, while many of the superintendents responsible for carrying out the appraisal that took part in their study perceived the process as providing helpful feedback, many school principals themselves did not, instead raising concerns about a predominantly summative, routine and meaningless exercise.

These findings indicate the need for a clear communication of the purposes of an appraisal process as well as the involvement of both evaluators and school leaders in the development of appraisal systems to ensure that the purposes are shared by everyone concerned. The studies by Thomas et al. (2000), Gaziel (2008) and Davis and Hensley (1999) also point towards the importance of including a strong formative dimension in the appraisal process, which school leaders valued very highly. This may be particularly important considering that appraisal always also involves the risk of increasing school leaders' workload and stress levels.

Procedures

This section describes countries' appraisal procedures. This includes the analysis of the frequency of appraisal, of the ways in which appraisal aspects and criteria are defined, of the reference standards that are used, of the value of professional school leadership standards and of the combination of instruments that evaluators use to gather information about a school leader's performance.

Frequency of school leader appraisal

The frequency of appraisal varies greatly across countries (Pont et al., 2008a), also depending on the functions of appraisal. Among countries which require the periodic appraisal of their school leaders, appraisal mostly takes place on an annual basis. This includes appraisal in various states and territories in Australia (e.g. Northern Territory), Chile (Performance Appraisal), various provinces and territories in Canada, Korea (School Principal Appraisal for Professional Development), Mexico (centrally managed schools at ISCED level 3), New Zealand, Slovenia, various autonomous communities in

Spain, and Northern Ireland in the United Kingdom. A few countries allow for longer intervals between appraisals ranging from three (e.g. France [ISCED levels 1, 2 and 3] and Israel [ISCED level 1 and partially ISCED level 2]), to four (Portugal) and five years (Belgium [French Community], some provinces and territories in Canada). Depending on the purposes of an appraisal scheme, appraisal may be additionally required before an employment-related decision (e.g. to inform contract renewal procedures). To give an example, in the Northern Territory in Australia, school principals' performance is reviewed twice within a one-year timeframe, but also six months before the end of a school principal's contract. Other systems are more flexible in their requirements and only specify a maximum period of time within which the responsible authorities need to carry out an appraisal (e.g. some states and territories in Australia, including Victoria, and Belgium [Flemish Community]). Legislation for Belgium (Flemish Community) specifies that appraisal needs to be conducted within a maximum period of four years. In Victoria, Australia, the Principal Class Performance and Development Plan can similarly cover a period of up to four years to accommodate individual and school needs. Among countries which require the appraisal of school leaders exclusively for employment-related decisions and contract renewal, the frequency of appraisal depends on the typical contract length for school leaders. In Poland and the Slovak Republic, school principals are appraised every five years, in the Czech Republic every six years. In countries where both school principals and deputy school principals are appraised, regulations for the frequency of appraisal are generally the same.

In a few countries, an additional appraisal can take place at the request of school leaders and/or evaluators or in the case of underperformance. In Ontario, Canada, for example, both evaluators and school leaders can request an additional appraisal in a non-appraisal year, which reinitiates the appraisal cycle. In France (ISCED levels 1, 2 and 3), the appraisal process can be additionally carried out at the request of a school principal that wishes to transfer to another school. In the case of underperformance identified as part of an appraisal, some countries require a compulsory appraisal to follow up on a school leader's weaknesses and to inform summative and/or formative consequences (e.g. some Australian states and territories, Belgium [French Community], some provinces and territories in Canada, Israel [ISCED level 1 and partially ISCED level 2], New Zealand, and Portugal).

The frequency and timing of appraisal are crucial to make appraisal a meaningful exercise at different stages of a school leader's career. For instance, appraisal may constitute an important lever for providing formative feedback to ease a school leader's transition into a new post or to maintain successful leadership over time (Mattson Almanzán et al., 2011). Unfortunately, not much research exists on the benefits and disadvantages of different approaches. Pont et al. (2008a) suggested an annual performance review period as most appropriate, but stressed the need to bear in mind the different functions of appraisal when setting requirements for the appraisal frequency. Following a review of current practices in school districts in the United States, Brown-Sims (2010) similarly recommended school leader appraisal on an annual basis. Lashway (2003) and Parylo et al. (2012b) emphasised the benefits of designing appraisal as a cyclical process that involves various meetings between school leaders and evaluators for appraisal (e.g. for preparation, data collection and follow-up) to function as a tool for the improvement of school leadership practices and behaviours and for integrating appraisal into the daily routine of school leaders' work. Piggot-Irvine (2003), however, also stressed the need to allow for sufficient time to carry out the appraisal to make it effective, to give school leaders room to improve and to decrease the risk of stress for those involved.

Appraisal aspects and criteria

School leader appraisal fundamentally relies on a shared conception of successful school leadership. Before appraising a school leader's performance, it needs to be clearly established what effective, successful or proficient leadership means (Duke, 1992). The way effective leadership is conceptualised forms the core of the expectations that school leaders need to satisfy and the aspects and criteria that are appraised, i.e. of the content and performance standards. In turn, through the definition of aspects and criteria, appraisal always communicates one vision of successful school leadership and the level of performance that can be achieved by the most effective school leaders. Developing a clear, research-based understanding of effective school leadership and building consensus about the meanings of good school leadership, therefore, constitutes one of the key challenges for school leader appraisal. For appraisal to be understood as fair and meaningful by school leaders, they need to be clear about the meanings of effective leadership. Role ambiguity resulting from a lack of clear expectations may lead to high levels of stress among school leaders (Pont et al., 2008a).

Considering appraisal's functions for development and accountability, the aspects and criteria that appraisal assesses must reflect the needs of individual school leaders, but be aligned with the wider goals of schools, local education systems and the education system as a whole at the same time (Davis and Hensley, 1999). A set of reference standards may help ensure a fair, valid and reliable appraisal process in relation to a shared conception of successful school leadership as well as overall education goals, local contexts and individual needs (also see Chapter 3). References for the definition of appraisal aspects and criteria typically comprise national education goals, professional school leadership standards, legislation on school leaders' roles and responsibilities, school development plans, school improvement plans and individual job descriptions.

The definition of appraisal aspects and criteria often follows broadly similar lines for school principals and deputy principals. Depending on the position (i.e. school principal, deputy school principal) and the school level (i.e. primary and secondary schools), different appraisal aspects and criteria may be used and emphasised and the appraisal of school principals may include a wider or narrower range of aspects and criteria. Differences in responsibilities between leadership positions are often reflected through reference documents such as individual job descriptions.

Many countries focus on the appraisal of school leaders' competences and related outcomes through the setting of individual goals and objectives

Depending on the ways in which the nature of leadership in schools is conceptualised, appraisal can focus on different types of aspects and criteria. This includes inputs, i.e. a school leader's personal traits; processes, i.e. a school leader's compliance with legal requirements or a school leader's performance of specific tasks and related competencies; and outcomes, i.e. the outcomes of a school leader's practices or wider school outcomes (Ginsberg and Berry, 1990; Duke, 1992). Within the governance framework and depending on the allocation of responsibilities for appraisal, aspects and criteria can be defined in various ways (Goldring et al., 2008). Central/state education authorities may provide a standardised set of responsibilities, tasks and competencies that are, then, assessed by the evaluators carrying out the appraisal (e.g. through a checklist or in an open format). Alternatively, central/state authorities may set broad guidelines and a list of responsibilities, tasks and competences from which evaluators and school leaders may choose areas to focus on that are relevant for their particular context. At times,

evaluators and school leaders may add additional aspects and criteria. Appraisal models may also leave it to the discretion of local or regional authorities and stakeholders to collaboratively define a set of content and performance standards.

Many countries follow a similar approach concerning the type of aspects and criteria that are assessed and the ways in which these are defined. Most often school leaders are assessed on the performance of their tasks and responsibilities and the overall quality of their practices and behaviours. In Belgium (Flemish Community), the selection of appraisal aspects and criteria is at the discretion of the evaluators. The policy framework in Belgium (French Community) defines three compulsory areas of competence (pedagogical leadership, interpersonal skills and resource management) that evaluators can complement with additional aspects and criteria. School principals in French primary schools are also assessed on their performance in four competency areas (general leadership, pedagogical leadership, community relations and resource management). In Portugal, the internal part of the appraisal process examines school leader's general leadership competencies, the efficacy, efficiency and quality of their practices and behaviours, and participation in professional development activities.

In a number of countries, the appraisal of school leaders' practices and behaviours includes the collaborative setting of individual objectives and an assessment of the extent to which these pre-determined goals have been met (e.g. various states and territories in Australia [e.g. Northern Territory and Victoria], various provinces and territories in Canada [e.g. Ontario], Chile [Performance Appraisal], France [ISCED levels 2 and 3], New Zealand and Portugal [internal part of the appraisal process]; also see Pont et al., 2008a). The objectives are often set in alignment with different reference standards (e.g. school development plans, job descriptions and professional standards) to ensure they reflect system, school and individual needs. When setting objectives, school leaders and evaluators typically identify a number of adequate information sources and instruments to gather relevant and valid information. On the basis of this information, school leaders are, then, appraised on the extent to which they have achieved their objectives. The appraisal of many school leaders is, therefore, outcomes- as well as process-based.

Examples of systems that combine the local definition of individual performance objectives with the use of a set of reference standards include various states and territories in Australia; Ontario, Canada; Chile (Performance Appraisal); France (ISCED levels 2 and 3); New Zealand; and Northern Ireland in the United Kingdom.

- In the majority of states and territories in Australia, school principals are appraised with reference to state standards for school leadership as well as a description of the general and professional duties of school leaders. In the state of Victoria, for example, school principals and deputy school principals define their learning needs as part of their annual performance and development cycle with reference to a state-wide developmental learning framework and the local school improvement plan. They, then, provide details of the leadership capabilities they intend to develop, indicate professional learning actions to build their capacity, and choose evidence they will use to monitor their growth and development.[23]
- In Ontario, Canada, appraisal is based on a performance plan that a school's board and school principal develop in collaboration. The aspects and criteria must be results- and goal-oriented and targeted towards the improvement of student achievement and well-being. The performance plan includes goals, strategies and actions, practices and competencies, as well as methods and indicators to measure the achievement of the pre-defined objectives. It is informed by personal

development goals as well as the board's and school's improvement plans, the school community and local context, and ministry priorities.[24]

- In Chile (Performance Appraisal), school principals in public schools are required to agree on two to four institutional targets and corresponding indicators as part of the regular mandatory appraisal process. These targets must be aligned with an overall institutional objective and seek to improve school practices and outcomes as defined through the ministry's School Management Quality Framework (*Modelo de Calidad de la Gestión Escolar*). In addition, school principals need to agree together with their evaluator on two to four professional development targets that are aligned with the Framework for Good School Leadership (*Marco para la Buena Dirección*). The achievement of both sets of targets, for which school principals need to provide evidence, is, then, given equal weight when forming the final appraisal result. The Chilean Ministry of Education has developed an online platform that provides guiding materials, space for discussion as well as a platform for appraisers and school principals to develop their target agreements and track and record them over time.[25]
- In France, school principals (*chefs d'établissement*) in public secondary schools are appraised against the meeting of objectives. School principals are required to provide their evaluators with a comprehensive analysis of the strengths and challenges of their schools (*diagnostic de l'établissement*) and a proposal of three areas for improvement in the medium term together with related performance indicators. In a subsequent meeting, the evaluator and the school principal establish a common framework and actions to achieve the objectives resulting in a mission statement and description of the general and professional duties of the school principal (*lettre de mission*) that is aligned with national education goals and strategies. The mission statement may be further modified at the request of the evaluator or school principal throughout the appraisal period. At the end of the appraisal period, the extent of achievement of the objectives is assessed through the preparation of a progress report by the school principal. In addition to these objectives, school principals are also appraised on the basis of an observation of their overall leadership (Dos Santos and Rakocevic, 2012).
- In New Zealand, appraisal by the school boards of trustees is based on an annual performance agreement that entails the identification and development of appropriate indicators. The performance agreement contains annual objectives for the school principal, including one or more professional development objectives, as well as the professional standards relevant to the role (e.g. school principal and teaching school principal). The professional standards provide a baseline for assessing satisfactory performance within each area of practice. The performance agreement must also reflect the school's strategic and annual plans, the school principal's job description and the New Zealand Teachers Council criteria for registration as a teacher.[26]
- In Northern Ireland in the United Kingdom, principals together with their evaluators establish a set of objectives at the beginning of the appraisal cycle for the year ahead. They reflect on possible outcomes and agree how best to keep progress. The objectives need to be related to three key areas: 1) leadership and management; 2) pupil and curriculum development; and 3) the personal and professional development of the principal. The objectives also need to reflect the School Development Plan. During the review year, related evidence is collected

and the progress towards the objectives is assessed in a final review discussion that results in a Review Statement (Department of Education, Northern Ireland, forthcoming).

In the case of the appraisal of deputy school principals and other school leaders, processes for the definition of performance objectives are often similar (e.g. various states and territories in Australia, including Victoria, various provinces and territories in Canada, Chile [Performance Appraisal], France [ISCED levels 2 and 3], and Northern Ireland in the United Kingdom). Chile and France provide two examples for the ways in which the setting of performance objectives for principals and deputy principals can be aligned. In Chile, the individual Performance Appraisal system requires technical-pedagogical leaders to set a similar set of objectives to those of school principals that are related to an institutional objective, institutional targets as well as professional development targets. The school-related targets must be linked to the school principal's institutional target and the extent to which teaching and technical-pedagogical leaders contribute to institutional targets is assessed. The targets for teaching and technical-pedagogical leaders are set in collaboration with school principals. Similarly in France, school principals themselves establish a mission statement (*lettre de mission*) for their deputy school principals in alignment with their own objectives and the responsibilities that are delegated to their deputy school principal.

Research increasingly suggests the benefits of focussing on school leaders' practices and behaviours, i.e. processes, and related outcomes. This reflects considerations about the extent to which aspects and criteria can be validly and reliably measured, the feasibility with which aspects and criteria can be assessed, and school leaders' control over the aspects and criteria they are judged on. Based on insights that leadership effects are mainly indirect, and that accountability for outcomes is important, but only in as far as these are reasonably related to a school leader's performance, Duke (1992) argued for the appraisal of leadership outcomes, i.e. of the desired consequences for specific leadership tasks, practices and behaviours. As he suggested, an emphasis on the direct consequences of a school leader's tasks, practices and behaviours may also support school leaders to prioritise their work and, therefore, help to decrease stress levels. The individual setting of objectives and the assessment of practices and behaviours seem to consider the control school leaders have over the aspects they are appraised against (Ginsberg and Thompson, 1992; Pashiardis and Brauckmann, 2008). An objectives- and task-based approach, furthermore, allows appraisal to focus on aspects most relevant to a particular school leader and school (Ginsberg and Berry, 1990). Scope for the local selection of appraisal aspects in line with central guidance and/or the collaborative setting of objectives at a local level may help make appraisal manageable and relevant for local contexts. It may allow evaluators and school leaders to focus on priority areas relevant to their context and analyse these in greater depth (Goldring et al., 2007).

A literature review on school principal appraisal practices and related research also stressed that school leaders should be involved in their appraisal through the definition of appraisal goals and objectives (Davis et al., 2011), as is a common practice in many countries. However, fairness concerns are equally important. Appraising school leaders against individually defined outcomes and objectives requires due attention to the contextual and organisational factors that influence whether outcomes and objectives can be reached. Duke (1992) highlighted that school outcomes can also be achieved despite a school leader's practices and behaviours. Even though the appraisal of processes or behaviours can be difficult, taking also processes or behaviours into account may enable evaluators to not only assess what leaders achieve, but how (Goldring et al., 2008).

School leader appraisal assesses a variety of aspects and criteria

The leadership practices and behaviours and related objectives and outcomes that appraisal is supposed to focus on as set out in central/state policy frameworks typically include:

- general leadership (e.g. leadership vision, values, ideals)
- pedagogical leadership (e.g. leading teaching and learning, setting learning objectives, curriculum development)
- organisation development (e.g. strategic planning, professional development of others)
- school climate (e.g. dealing with diversity, school violence and bullying)
- community relations (e.g. work with families, communities, external partners)
- evaluation and accountability (e.g. skills for school self-evaluation, teacher appraisal, student assessment)
- resource management (e.g. administration of buildings, financial and human resources, administration, time resources of staff)
- interpersonal skills (e.g. communication, interaction, conflict management, professional dialogue).

Policy frameworks in different countries emphasise different aspects with varying scope for pedagogical leadership (see Table 7.A.1). Many countries' appraisal systems suggest and/or require evaluators to include pedagogical leadership as one of the appraisal aspects (e.g. many states and territories in Australia, Belgium [French Community], many provinces and territories in Canada, Chile [Performance Appraisal], France [ISCED levels 1, 2 and 3], Israel [ISCED level 1 and partially ISCED level 2], Korea [School Principal Appraisal for Professional Development], New Zealand, Poland, Slovenia, various autonomous communities in Spain, and Northern Ireland in the United Kingdom). To give two examples, in Chile, at least two of the school principals' institutional targets must be related to the ministry's School Management Quality Framework and at least one of these needs to refer to the domain of results, a further one to the areas of leadership or curriculum management. The institutional targets of technical-pedagogical leaders are related to the institutional targets of school principals and need to fulfil the same requirements, except for the requirement to set at least one target related to results. The Korean School Principal Appraisal for Professional Development also requires a strong focus on pedagogical leadership. Evaluators assess school principals on their school education plans in terms of their management of school operation goals (i.e. their formulation and operation of the curriculum, development of creativity and character in students), school principals' school supervision (i.e. classroom instruction improvement, autonomous supervision activities) and school principals' teacher management. The Korean model also includes school principals' management of facilities and budgets (Kim et al., 2010). Only a few countries include responsibilities and competences related to evaluation and accountability (e.g. many states and territories in Australia, Israel [ISCED level 1 and partially ISCED level 2], Mexico [centrally managed schools at ISCED level 3] and various autonomous communities in Spain). Only two countries participating in the OECD Review reported that appraisal frameworks require the evaluation of practices and behaviours related to both, pedagogical leadership, and

evaluation and accountability (various states and territories in Australia and various autonomous communities in Spain).

However, when considering appraisal aspects and criteria, it is essential to bear in mind that the actual implementation of appraisal procedures is key and that the actual focus of appraisal may differ from requirements and general guidelines as set out in central/state policy frameworks (Kimball et al., 2009). Some of the stakeholders interviewed during the OECD Country Review of New Zealand as part of the OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes, for example, raised concerns about a compliance-driven process that focuses too much on budget and finance issues despite guidelines that suggest a focus on pedagogical leadership (Nusche et al., 2012).

Focussing on pedagogical leadership may be particularly effective for improving teaching and learning

The leadership practices and behaviours that appraisal in different contexts evaluates need to be contextualised with research on effective school leadership as well as school leaders' current practices and behaviours.

Research has built up a strong knowledge base of successful school leadership and the ways in which school leaders can influence teaching and learning through an overall focus on pedagogical leadership, thus strengthening the case to include a strong dimension of pedagogical leadership as part of the appraisal aspects and criteria (see section on the analytical approach for this chapter). A focus on pedagogical leadership as part of school leader appraisal also seems essential within a wider evaluation and assessment framework. In more than half of the countries that took part in TALIS, schools with a more pronounced pedagogical leadership style tended to link teacher appraisal with teachers' participation in professional development. In many TALIS countries, schools in which directors focussed on pedagogical leadership were also more likely to take account of innovative teaching practices in the appraisal of their teachers (OECD, 2009).

As research has pointed out, school leaders' responsibilities for evaluation, assessment and accountability, and related practices and behaviours that focus school leaders' attention towards teaching and learning (e.g. through classroom observation and teacher feedback) are also closely linked to pedagogical leadership (The Wallace Foundation, 2012). Research by Ovando and Ramirez Jr. (2007), for example, demonstrated these links between pedagogical leadership and evaluation, assessment and accountability in their study on the relationship between pedagogical leadership and teacher appraisal. As their research indicated, depending on a school principal's leadership style, school principals can use teacher appraisal as one tool to improve instruction and student learning. Research indicated that school leaders that focus on goal-setting, monitoring, assessment and evaluation can also positively influence teacher performance and learning environments (Pont et al., 2008a). Researchers have, furthermore, stressed the need to select, prioritise and weight a select number of multiple appraisal aspects that are precise, achievable, and measurable considering school leaders' vast range of tasks and responsibilities (Goldring et al., 2008; Reeves, 2009; Clifford and Ross, 2011). As Portin et al. (2006) and Davis et al. (2011), therefore, suggested, focussing on a few practices that have been shown to have a potentially high impact on schools may constitute one strategy to make appraisal more effective. Considering the role of pedagogical leadership for improving teaching and learning and its

interrelationship with responsibilities for evaluation, assessment and accountability, there are arguments for focussing appraisal on both of these domains.

Only little research has been done on the ways in which the focus of appraisal influences actual school leadership practices. However, two recent studies point towards the potential impact of an appraisal focus on pedagogical leadership. Sun and Youngs's (2009) study of the effects of school principal appraisal on leadership practices and behaviours in school districts in Michigan highlighted the importance of the appraisal aspects and criteria that are appraised. Their study, furthermore, suggested that, although not in a causal relationship, school principals were more likely to engage in pedagogical leadership practices and behaviours when appraisal focussed on school goal setting, curriculum design, teacher professional development and appraisal, and the monitoring of student learning. Subsequent research by Sun et al. (2012) on appraisal in Michigan, United States, and Beijing, China, similarly suggested that appraisal practices that included a focus on curriculum, instruction and student learning encouraged school principals to exercise influence in these domains.

School leaders make use of a variety of leadership styles

Research on actual school leadership responsibilities and practices and behaviours provides a further perspective on current approaches to appraisal aspects and criteria and an interesting context to the available research on the nature of effective leadership in schools.

Roles and responsibilities of school leaders differ among countries as well as between formal school leadership positions and school levels. In some systems, school principals' tasks may focus on ensuring coherence between teacher development and the educational goals of a school, in other systems, school principals may focus more on supervising classroom instruction, for example (OECD, 2010). Such role differences are essential for the expectations that school leaders are judged against. The OECD project on Improving School Leadership also found that across OECD countries, the responsibilities of school principals and their autonomy in school affairs vary considerably. In many countries, school leaders have traditionally held a more administrative role geared towards management, administration and the translation of policies into schools and classrooms (Pont et al., 2008a). In many contexts, job descriptions are still not clearly tailored towards pedagogical leadership practices that research suggests as the core of effective school leadership, but focus instead on the traditional tasks of school principals as managers and administrators (Pont et al., 2008a).

International surveys, even though based on school principals' own responses and their interpretation of the questions asked, provide insights into actual school leadership practices and behaviours.

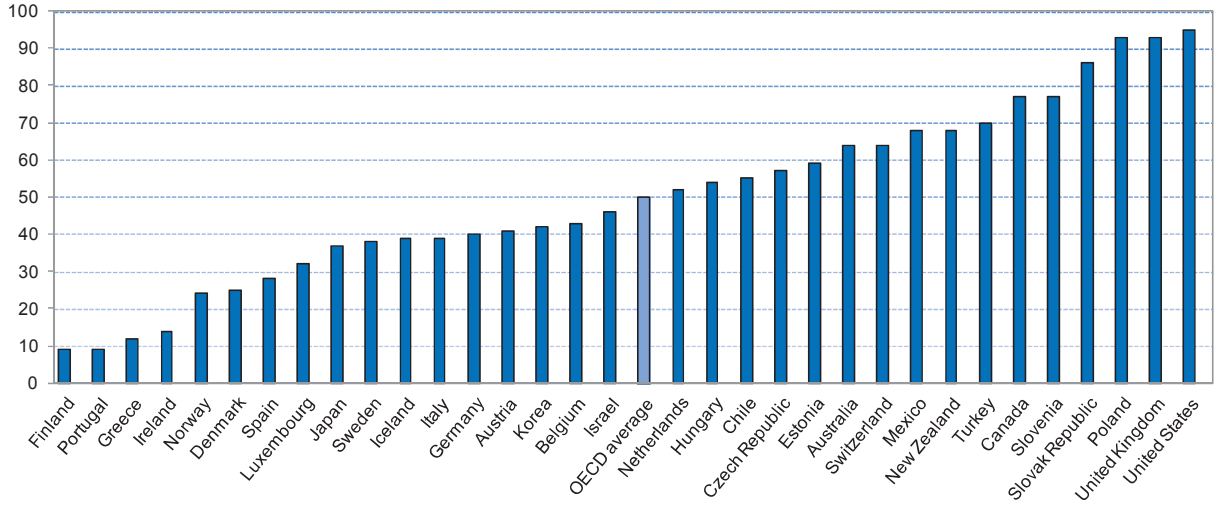
- A survey conducted among school leaders as part of TALIS gives information on the practices and behaviours of lower secondary school principals across 23 countries (OECD, 2009). The TALIS survey distinguished between an instructional leadership style and an administrative leadership style. Based on school principals' responses, TALIS found that, out of the 23 participating countries, school principals in 10 countries made greater use of pedagogical leadership, with a sizeable group of school principals combining pedagogical with administrative leadership. While arguing that effective school principals are likely to display both leadership styles, TALIS concluded that the pedagogical leadership paradigm has made some progress in all participating countries, albeit

in some much more than in others. Even though different leadership styles often co-exist, and a trend towards pedagogical leadership can be observed in many contexts, however, more administrative leadership styles are still predominant in various contexts (OECD, 2009).

- Data on school principals' leadership collected through the OECD Programme for International Student Assessment (PISA) sought to unravel the extent to which school principals are active in improving teaching practices and the working environment within schools. The principal survey administered through PISA found that, on average among OECD countries, 93% of students attended schools whose school principals reported that they ensure that teachers' work reflects the school's educational goals quite often or very often. Over 86% of students attended schools whose school principals reported to quite often or very often take the initiative to discuss a problem teachers may have in their classrooms. 75% of students attended schools whose school principals reported to quite often or very often use student performance results to develop the school's educational goals. However, only 61% of students attended schools whose school principals reported to quite often or very often consider exam results when making decisions regarding curriculum development and only half of students attended schools whose school principals reported to quite often or very often observe classes (OECD, 2010, see Figures 7.4 and 7.5). PISA, thus, also provided a mixed picture of current school leadership practices and behaviours.

Figure 7.4 School principals observing instruction in classrooms (2009)

Percentage of students in schools whose principals reported that they had observed instruction in classrooms "quite often" or "very often" during the last school year (%)

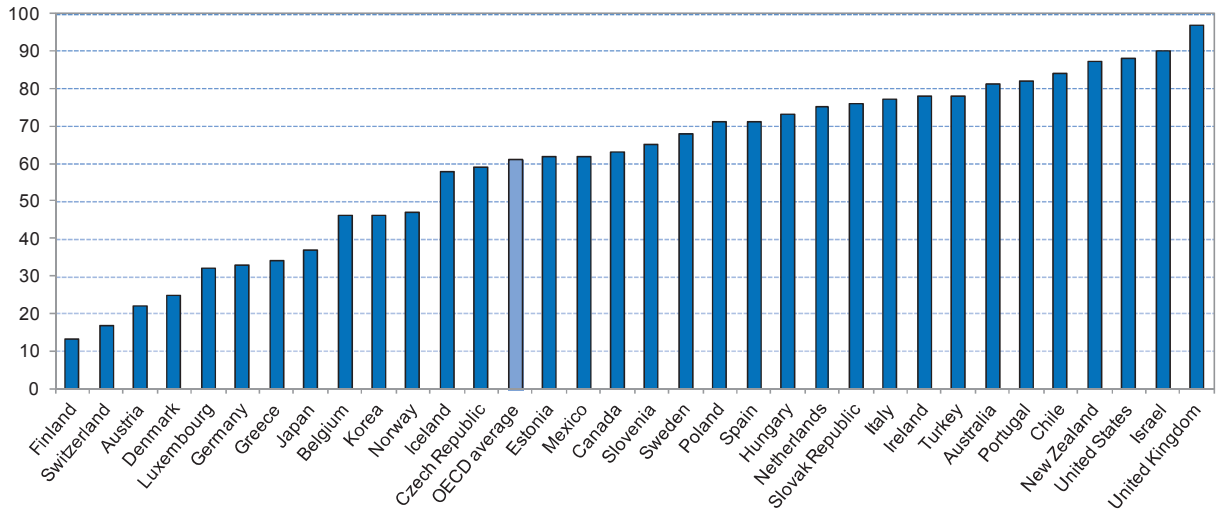


StatLink http://dx.doi.org/10.1787/888932791533

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

Figure 7.5 School principals using exam results for curriculum development (2009)

Percentage of students in schools whose principals reported that they had observed instruction in classrooms "quite often" or "very often" during the last school year (%)



StatLink http://dx.doi.org/10.1787/888932791552

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

However, school leaders also face challenges to move towards a more pedagogical leadership style. In systems with flat hierarchical structures, school leaders with ambitions for pedagogical leadership that involves classroom observation, for example, may encounter resistance from teachers. In such systems, school leaders may be hesitant to take responsibility for pedagogical leadership (Pont et al., 2008a). The Country Background Reports of some of the countries participating in the OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes equally stressed challenges for school leaders to move towards a focus on pedagogical leadership practices and behaviours. In the Czech Republic, the often large administrative burden in basic schools that offer primary and lower secondary education makes it difficult for head teachers to manage educational processes and outcomes. Instead, most of their work is often of an administrative, management or budgetary nature. Despite a high degree of autonomy to manage teachers, many Czech head teachers lack preparation for their responsibilities, particularly in the area of leading teaching and learning. Professional development opportunities often also lack a focus on pedagogical leadership. (IIE, 2011). Countries as diverse as Denmark, New Zealand, Norway, Portugal and Slovenia similarly reported that school principals most often focussed on administrative tasks and could focus more on their pedagogical leadership or that school leaders faced difficulties in reconciling the need to carry out administrative, managerial and leadership tasks and responsibilities (Danish Ministry of Education and Rambøll, 2011; New Zealand Ministry of Education, 2010; Norwegian Directorate for Education and Training, 2011; Portuguese Ministry of Education and Science, forthcoming; Brejc et al., 2011). The OECD Country Review of Mexico also noted that, despite the set of desirable responsibilities, tasks and competencies communicated through central education authorities, school leaders' are often preoccupied with the administration of schools (e.g. ensuring a safe infrastructure and compliance with legislation), leaving often little time for focussing on pedagogical leadership tasks that would have a great effect on the quality of education in their school (Santiago et al., 2012b).

Policy makers face the challenge of ensuring a focus on pedagogical leadership and scope for local contextualisation

These insights into school leaders' roles and responsibilities and actual practices and behaviours suggest that, while school leaders are moving towards a greater focus on pedagogical leadership, greater efforts are needed to support school leaders to incorporate more pedagogical leadership practices and behaviours into their work and to balance management, administrative and pedagogical leadership tasks and responsibilities effectively. Limited research indicates that appraisal can support leaders to strengthen their pedagogical leadership (Sun and Youngs, 2009; Parylo et al., 2012b; Sun et al., 2012). However, policy frameworks for school leader appraisal in a few contexts and appraisal as it is implemented in some countries do not focus sufficiently on pedagogical leadership. Policy makers, therefore, face the challenge of ensuring that appraisal is focussed on the ultimate objective of improving teaching and learning for all students and that appraisal is designed to strengthen pedagogical leadership in schools. This seems particularly relevant considering the need to focus and weight a limited number of appraisal aspects and criteria in light of the range of school leaders' tasks and responsibilities.

However, within an overall focus on pedagogical leadership, policy makers need to pay due attention to the contextual nature of successful school leadership. While recent research has built up a strong knowledge base of the core of pedagogical leadership and highlighted the benefits of focussing on pedagogical leadership as part of school leader appraisal, research has at the same time stressed the need for contextualisation (also see the second section on the analytical approach on which this chapter is based). School leadership is a complex task that often requires spontaneity and prioritisation of tasks according to circumstances and adaptation to school-level factors and characteristics. Heck and Marcoulides (1992) have, therefore, suggested to also consider the ways in which school leaders adapt their practices and behaviours to the particular context in which they work as part of their appraisal.

Research in the United States has developed a few models that provide some indications for ways in which appraisal can focus particularly on pedagogical leadership. Based on a conceptual framework for effective pedagogical leadership (Murphy et al., 2006) and a conceptual framework for appraisal (Porter et al., 2006), Goldring et al. (2007) proposed one appraisal model for school leaders. Defining leadership behaviour as the interrelationship of core components and core processes, they suggested the appraisal of a set of functional competencies (high standards of student performance, rigorous curriculum [content], quality instruction [pedagogy], culture of learning and professional behaviour, connections to external communities and systematic performance accountability) and behavioural competencies (planning, implementing, supporting, advocating, communicating and monitoring). Reeves (2009) provided an example of a selection of leadership practices and behaviours that can form part of appraisal that focuses on pedagogical leadership. These include resilience; personal behaviour; student achievement; decision making; communication; faculty development; leadership development; time, task and project management; technology; and learning.

Appraisal requires the specification of a number of clearly differentiated performance levels

Appraisal typically relies on the specification of levels of performance for each of the aspects that are appraised, i.e. specifications of what counts as effective leadership in each of the appraisal aspects or the level of acceptable performance when reaching

pre-determined objectives (Ingvarson et al., 2006). The specification of performance levels describes a qualitative dimension for appraisal aspects that goes beyond considerations whether a competency, practice or behaviour has been demonstrated or an objective has been met, but considers how, and to what extent (Duke, 1992).

While some countries do not give school leaders a rating or leave it to the discretion of local authorities and evaluators to establish a satisfactory level of performance (e.g. Belgium [French Community], various provinces and territories in Canada, New Zealand, various autonomous communities in Spain, and Northern Ireland in the United Kingdom) or rely on a binary scale to rate school leaders (e.g. satisfactory/unsatisfactory, pass/fail) (e.g. Belgium [Flemish Community] and various provinces and territories in Canada), most countries award a rating with a number of performance levels. In these contexts, the number of performance levels ranges between three (e.g. Poland), four (e.g. Chile [Performance Appraisal], France [ISCED levels 1, 2 and 3] and Mexico [centrally managed schools at ISCED level 3]) and five (e.g. Israel [ISCED level 1 and partially ISCED level 2], Korea [Evaluation for School Management and School Principal Appraisal for Professional Development], and Portugal[27]). The Slovenian appraisal system relies on a sliding scale from 0 to 100. In Chile, the Performance Appraisal system similarly relies on a percentage scale with four related levels of performance. School leaders and school leadership teams need to reach a minimum threshold of 50 to be rated as satisfactory.

Recent research has emphasised the benefits of clearly distinguished and transparent appraisal criteria, i.e. specified levels of performance for appraisal aspects, that are clearly communicated to and shared by school leaders and evaluators and consistently applied (Ginsberg and Berry, 1990; Reeves, 2009; Clifford and Ross, 2011). As Reeves (2009) highlighted, an appraisal scheme that lacks a clear differentiation between levels of performance risks ranking leaders as uniformly outstanding in all of the aspects that are appraised, therefore limiting the usefulness of appraisal (Reeves, 2009). To allow for sufficient differentiation between various levels of performance, Reeves (2009) proposed the introduction of four rating categories (exemplary, proficient, progressing, not meeting standards).

Various other researchers have stressed the need to take a school leader's background and experience into account when applying different performance levels (Pashiardis and Brauckman, 2008; Brown-Sims, 2010; Clifford and Ross, 2011). Ensuring scope to adjust appraisal criteria to a school leader's experience in general as well as the length of time a leader has spent in a particular school may improve the appraisal process by paying due attention to the impact that can be expected of a particular school leader (Goldring et al., 2007). As Goldring et al. (2007) stressed, while contextual variables and a school leader's experience should not be used as an excuse for underperformance, evaluators need to take them into account when interpreting school leaders' achievements, practices and behaviours.

In light of these considerations, policy makers need to find ways to establish clearly distinguished levels of performance that are consistently applied by competent evaluators, while also being sensitive to particular contexts and a school leader's background and experience.

An increasing number of countries have been developing professional standards for school leadership

An increasing number of systems have been developing professional standards or frameworks for school leadership on a country and/or state-wide level in an effort to define the nature and scope of successful school leadership and the expectations school leaders are

expected to fulfil. Box 7.8 provides two examples for the process of developing professional standards in the federal systems of Australia and the United States.

Box 7.8 Developing professional school leadership standards in federally governed education systems

Australia

All in all, around fifty sets of leadership standards and capability frameworks have been designed and developed in Australia. Recently, the Australian Institute for Teaching and School Leadership (AITSL) developed a national set of professional standards for school principals to provide a national framework for use in all of the country's schools and education systems as a step towards a more consistent nationwide approach. The National Professional Standard for Principals aims to define the role of school principals and to unify the profession nationally, to describe the professional practice of school principals in a common language and to make explicit the role of quality school leadership for improving learning outcomes. The national content standard is designed to assist in attracting, developing and supporting aspiring and practising school principals. It is intended to inform professional learning, to guide self-reflection, self-improvement and development, and to guide the management of self and others. The development of the standard lasted from early 2010 to mid-2011 and involved research, drafting, critical review, feedback and testing through a series of pilot studies, guided and supported by an external expert steering group. The group included expert representatives from state and territory government employers, the catholic and independent school sectors, national professional associations and education boards, leadership institutes and Commonwealth and national school leadership experts. Pilot studies tested a draft version of the standard from early February to May 2011 to examine authenticity, usefulness and the added value. The pilots also informed the final formulation of the standard as well as the implementation strategy.

United States

In the United States, the Council of Chief State School Officers (CCSSO) and the National Policy Board for Educational Administration (NPBEA) have developed and adopted a set of Educational Leadership Policy Standards: Interstate School Leaders Licensure Consortium (ISLLC) 2008 that are open for voluntary use and adaptation by individual states. This set of policy standards constitutes an update of an earlier set of standards, the 1996 ISLLC Standards for School Leaders, to reflect new knowledge about educational leadership and the changing policy context of education in the United States. The 2008 set of policy standards intends to strengthen the message that school leaders' primary responsibility is to improve teaching and learning for all children. As an explicit set of policy standards, the ISLLC 2008 Standards are designed to support policy makers, especially at the state level, to reflect about their leadership development system as a whole and its individual elements (e.g. preparation, licensing, induction, professional development and appraisal) rather than individual school leaders. As a basis for performance appraisal measures, the 2008 ISLLC standards can help states, district and schools to formalise expectations of school leaders and help create an aligned performance-based system overall.

The new standards flow from a two year revision process led by NPBEA in consultation with policy-oriented, practitioner based organisations, researchers, higher education officials and school leaders. A panel of researchers and school leadership experts identified the research base for updating the ISLLC 1996. The development of the standards was organised by a steering committee in several phases. All NPBEA member organisations identified a strategy to ensure membership input into the revision process. A first draft of the revised standards was, then, distributed among member and professional organisations and the research panel to gather feedback. Based on this process of input and feedback, the steering committee revised its drafts and finalised the 2008 set of policy standards.

Sources: Australian Institute for Teaching and School Leadership (AITSL), National Professional Standard for Principals website, see *www.aitsl.edu.au/school-leaders/national-professional-standard-for-principals/national-professional-standard-for-principals.html*; Council of Chief State School Officers (CCSSO), Educational Leadership Policy Standards: ISLLC 2008, see *www.ccsso.org/Resources/Publications/Educational_Leadership_Policy_Standards_ISLLC_2008_as_Adopted_by_the_National_Policy_Board_for_Educational_Administration.html*.

Professional standards differ in their complexity and specificity, but are typically structured around key dimensions of leadership that are broadly described, followed by a list of core functional and behavioural competencies that define how to put these into practice (Ingvarson et al., 2006; CEPPE, forthcoming). They may differentiate between various levels of experience as well as between school levels (i.e. primary and secondary schools). Box 7.9 provides three examples that illustrate the different ways in which professional standards have been defined.

Professional standards or leadership frameworks can provide a reference point for the development of a school leader throughout his or her career (e.g. for recruitment, professional development and career advancement). A number of appraisal frameworks rely on the use of professional school leadership standards as a reference point for appraisal aspects and criteria or individual performance objectives (see Table 7.A.1). This includes many states and territories in Australia, Belgium (French Community), various provinces and territories in Canada, Chile (Individual Performance Appraisal), Israel [ISCED level 1 and partially ISCED level 2], Mexico (centrally managed schools at ISCED level 3) and New Zealand.

A number of other countries, including Belgium (Flemish Community), France (ISCED levels 1, 2 and 3), Korea (School Principal Appraisal for Professional Development and Evaluation for School Management), Poland, Portugal, Slovenia, various autonomous communities in Spain, and Northern Ireland in the United Kingdom have appraisal systems in place that do not make use of professional standards or leadership frameworks.

Box 7.9 Different approaches to the definition of professional school leadership standards

Victoria, Australia

The state of Victoria, Australia, has developed a Developmental Learning Framework for School Leaders, as a fundamental element of its 2006 Learning to Lead Effective Schools strategy. The framework is intended to strengthen school principals' and teachers' leadership capacity. It can be used in various ways, e.g. for self-assessment, performance and development reviews, school leader selection, coaching and mentoring and leadership induction and planning. The Victorian leadership framework breaks new ground in being applicable to leadership throughout the school at all levels in the school, showing where a teacher or school leader is located on a leadership continuum and what they need to know and be able to do in order to improve. As such, the Victorian framework is based on the core belief that leadership is learnable. The framework describes development within five leadership domains: (1) technical; (2) human; (3) educational; (4) symbolic; and (5) cultural. Within each of these leadership domains, the framework lays out typically five progressive levels of competence and related capabilities. It defines what effective leadership looks like in practice at each of the different stages of development and growth and provides a clear direction about what it means to develop as a leader.

Sources: Victoria Department of Education and Early Childhood Development, *www.eduweb.vic.gov.au/edulibrary/public/teachlearn/leader/developmental_learning_framework_20070418.pdf;* Bastow Institute of Educational Leadership, *www.bastow.vic.edu.au/leadership-development/pages/school-leadership.aspx.*

Box 7.9 Different approaches to the definition of professional school leadership standards (*continued*)

Chile

Chile's Framework for Good School Leadership (*Marco para la Buena Dirección*) specifies criteria of effective school leadership that form the basis for professional development and performance appraisal. It aims to support the development of pedagogical leadership to respond to political, economic and social changes (e.g. Chile's national development strategy, decentralisation in education, democratisation, and globalisation). Based on stakeholder consultations and national as well as international experiences in school leadership, the Chilean school leadership model defines successful leadership as practices related to pedagogical, administrative and financial management. Accordingly, the Framework for Good School Leadership defines four areas of practice: (1) leadership; (2) managing the curriculum; (3) managing resources; and (4) managing the school climate, which are, then, defined in greater detail. In addition, Chile has a Good School Management Framework (*Modelo de Calidad de la Gestión Escolar*) in place. These guidelines similarly define successful school management as a set of processes related to leadership, managing the curriculum, school climate and student support, and managing resources, as well as results.[28]

Sources: SIMCE website (System to Evaluate the Quality of Education), see *www.simce.cl/fileadmin/Documentos_y_archivos_SIMCE/Material_de_apoyo_establecimientos/diagramaciondef2005BPDF.pdf*, *www.simce.cl/fileadmin/Documentos_y_archivos_SIMCE/Material_de_apoyo_establecimientos/ModelodeCalidaddelSACGE.pdf*.

New Zealand

New Zealand has developed a Kiwi Leadership for Principals (KLP) model that provides a statement of the expectations of school principals. Built on a core conceptualisation of educational leadership and stressing the need of building effective relationships as well as school leaders' attention to their particular contexts, KLP defines Leading Change and Problem-Solving as the two key leadership areas for school principals. The KLP model, further, identifies four areas of practice (culture; pedagogy; system; partnerships and networks) to reach these two objectives. Four educational leadership qualities underpin school leaders' ability to lead their schools: *manaakitanga* (leading with moral purpose), *pono* (having self belief), *ako* (being a learner), and *awhinatanga* (guiding and supporting). In alignment with this leadership framework, two sets of professional standards for primary and secondary school principals provide a baseline for assessing satisfactory performance within each area of practice (culture; pedagogy; system; partnerships and networks). New Zealand has been in the process of developing two further parts of the overall leadership strategy: Kiwi Leadership for Senior and Middle Leaders and Leadership for Māori-medium Leaders.

Source: New Zealand Ministry of Education website, Kiwi leadership for principals: principals as educational leaders, see *www.educationalleaders.govt.nz/Leadership-development/Professional-information/Kiwi-leadership-for-principals*.

Professional standards can help provide a clear description and create a shared understanding of the core practices of effective school leadership

Appraisal requires clear and realistic performance expectations to ensure that appraisal is feasible, fair, valid and reliable, and that it is geared towards improving student outcomes. Depending on how they are designed and implemented, professional standards for school leaders can help to clearly communicate such expectations (Pont et al., 2008a). Based on a literature review of professional standards for school leadership, Ingvarson et *al.* (2006) argued that professional standards may also help set reasonable boundaries to the scope of school leaders' work and realistic expectations.

Only limited research has been undertaken on the ways in which the use of professional standards affects school leadership practices, school outcomes and school leader appraisal (Kimball et al., 2009). Theoretically, however, professional standards, if

applied coherently, can contribute to a fair, valid and reliable appraisal process by ensuring that school leaders are appraised against an absolute standard of high performance and by ensuring that appraisers hold a shared conception of effective school leadership. Reference to a set of professional standards that clearly lay out what the best school leaders can achieve, furthermore, can help focus appraisal towards the improvement of all school leaders, especially those that perform well, but could aim still higher (Reeves, 2009). Kimball et al. (2009) conducted an empirical study on school principals' perceptions of an appraisal process that was standardised across a school district in the United States on the basis of a randomised trial. They sought to shed light on the ways in which a standards-based appraisal procedure affects the communication of performance expectations, the provision of performance feedback, perceived utility in improving performance, and school principals' satisfaction and perception of fairness. Although the survey results were only statistically significant for the aspect of higher feedback quality, linked to challenges for the implementation of the standardised procedures, the qualitative results suggested that principals appraised through the standardised procedures experienced better feedback and greater overall satisfaction with the process. The absence of research-based central/state professional standards or leadership frameworks may, then, weaken capacity for effective appraisal across a whole education system, especially where appraisal entails a large degree of local autonomy. However, Kimball et al. (2009) also highlighted that the effectiveness of a standards-based approach depends to a great extent on the way procedures are effectively applied by competent and trained evaluators.

Instruments and sources of information

A number of instruments and information sources are typically used for the appraisal of school leaders, including interviews, observations and school visits, self-appraisal and leadership portfolios, tests, stakeholder surveys and questionnaires, school climate surveys and student data. The choice of instruments and sources of information also depends on the type of appraisal aspects and criteria that an appraisal scheme focuses on (e.g. traits, compliance, competences and outcomes). While in most countries appraisal is based on the interaction between evaluators and school leaders, e.g. through interviews, meetings and school visits, a few countries complement the information gathered this way with further, more outcomes-oriented data, such as leadership portfolios, stakeholder surveys and information on student achievement. Depending on the policy framework for appraisal and the overall allocation of responsibilities, evaluators and school leaders themselves hold greater or lesser autonomy in determining the tools to collect information. In various states and territories in Australia, Belgium (French Community), in various provinces and territories in Canada, Chile (Performance Appraisal) and New Zealand, for example, it is at the discretion of school leaders and evaluators to select instruments and tools to gather information depending on the chosen appraisal aspects and criteria and/or individually defined performance objectives. The instruments and sources of information used to gather evidence about a school leader's performance may also differ between leadership positions, school principals and deputy school principals.

Policy makers need to bear in mind that the instruments and information sources that are used need to yield an accurate, fair, valid and reliable picture of a school leader's performance. As some research has pointed out, gathering information from multiple sources of evidence about a school leader's practices and behaviours constitutes one strategy to increase the accuracy, fairness, validity and reliability of an appraisal and help take into account the complexity of successful school leadership from a range of different

perspectives when judging a school leader's performance (Piggot-Irvine, 2003; Pashiardis and Brauckman, 2008; The Wallace Foundation, 2009; Brown-Sims, 2010; Leon et al., 2011; Sun et al., 2012). While the costs and efforts required need to be weighed with the benefits of using multiple sources of information, it is essential to gather a feasible amount of accurate, valid, reliable and useful information (Glasman and Glasman, 2010).

Interviews, school visits and observations

Meetings and discussions with school leaders constitute an essential way of gathering information about a school leader's performance. Such meetings between appraiser and appraisee may simply rely on a checklist that examines whether an appraisal aspect has been demonstrated (Lashway, 2003). The interaction between evaluators and school leaders may extend to observations of school leaders in their daily work, classroom observations and interviews with stakeholders (e.g. teachers and students) to gain an impression of the school climate.

A few countries base their appraisal system solely on a school visit and an interview between school leaders and evaluators (e.g. Belgium [Flemish Community], various provinces and territories in Canada, France [ISCED level 1], and Northern Ireland in the United Kingdom). Other systems, such as various states and territories in Australia, various provinces and territories in Canada, Chile (Performance Appraisal), France [ISCED levels 2 and 3], Israel [ISCED level 1 and partially ISCED level 2], Korea (Evaluation for School Management and School Principal Appraisal for Professional Development), New Zealand, Poland, Portugal, Slovenia and various autonomous communities in Spain complement interviews and meetings with further tools to gather additional information.

Within the overall timeframe for the frequency of appraisal that is determined by central/state regulations, meetings between school leaders and evaluators and/or school visits and observations may take place once or various times. Depending on the context, this is also at times defined through the central/state framework. Victoria, Australia, for example, requires annual review meetings within the timeframe that school leaders and evaluators select for appraisal. In Ontario, Canada, school leaders and evaluators must meet at least three times within the evaluation year designated as part of the five-year evaluation cycle to facilitate professional dialogue between appraiser and the teaching professional being appraised.

As Duke (1992) pointed out, the frequency with which appraisal meetings take place imply different messages about the focus of appraisal. A continuous appraisal process communicates a more formative image based on the idea that a school leader's performance can be improved. An appraisal process that only requires one meeting between school leaders and evaluators rather places emphasis on the question whether a result has been achieved or a skill has been demonstrated. Based on an extensive review of literature on school leader appraisal, Leon et al. (2011), therefore, highlighted the opportunity for ongoing and regular interaction between evaluators and school leaders, e.g. through interviews and school visits, as one of the key features of effective appraisal. A high frequency of meetings and school visits may increase the effectiveness of these tools to gain a better picture of a school leader's performance and produce valid information (Pashiardis and Brauckman, 2008). However, the time and efforts required for the longer-term observation of school leaders also involves costs that need to be considered (Ginsberg and Berry, 1990).

Self-appraisal and leadership portfolios

Self-appraisal constitutes another frequently used instrument, often in relation to individually set objectives. Self-appraisal may take the form of a self-reflection journal or the compilation of a longer-term leadership portfolio. When compiling a portfolio, school leaders typically collect evidence to demonstrate progress towards pre-defined objectives. Evidence may include, among others, parent newsletters, staff meeting notes, photographs of classroom activities and records of dropout rates (Pashiardis and Brauckman, 2008). Reflective conversations with colleagues constitute a further form of self-appraisal (Portin et al., 2006).

In various countries appraisal may require school leaders to reflect on their own performance. In various Australian states and territories, various provinces and territories in Canada, New Zealand and Portugal appraisal may include self-appraisal as part of the process. In Israel (ISCED level 1 and partially ISCED level 2) both school principals and deputy school principals need to appraise their own performance, but only school principals are required to prepare a portfolio. School leaders in Chile (Performance Appraisal) and Slovenia and various autonomous communities in Spain are also required to collect and present evidence in the form of a portfolio.

The use of self-appraisal and portfolios entails both benefits and drawbacks that policy makers need to weigh. Scope for school leaders to express their own views about their performance is generally considered good practice in school leader appraisal (Pashiardis and Brauckman, 2008; Clifford and Ross, 2011). Self-appraisal and portfolios can provide such a platform by fulfilling a twofold function. Firstly, self-appraisal and portfolios can provide rich information with leaders themselves demonstrating their capabilities and performance. As Lashway (2003) stressed, portfolios can be used in a very flexible way and according to the needs of a particular school leader. Secondly, school leaders' involvement through self-appraisal or portfolios provides an opportunity for professional reflection, whether over a school year or their whole leadership career. Allowing school leaders to appraise themselves and/or to prepare a portfolio may also help create trust and a positive relationship between school leaders and evaluators (Pashiardis and Brauckman, 2008).

However, validity and reliability concerns as well as the costs of self-appraisal and/or portfolios in terms of the required workload for school leaders also need to be taken into account. Considering the subjective involvement of leaders in their own appraisal, self-appraisal and portfolios may be more suitable as tools for improvement through reflection. Other sources of information may usefully complement the information and perspectives gathered by school leaders themselves (Goldring et al., 2008). Clear guidelines and criteria on self-appraisal and the preparation of portfolios may increase validity and reliability and ensure that portfolios are used in a meaningful way (Ginsberg and Berry, 1990; Lashway, 2003). Self-appraisal and/or the preparation of portfolios may also require considerable time and efforts from school leaders. Bearing in mind the often heavy workload and range of responsibilities of school leaders, this is an important point to consider (Ginsberg and Berry, 1990).

Johnston and Thomas (2005) carried out research on the experiences of school principals with a pilot project that was part of the preparations for the introduction of a state-wide portfolio appraisal system for newly appointed school principals in Ohio, United States.[29] Their research aimed to provide insights into the perceived benefits and drawbacks of the use of portfolios. Opinions about the effectiveness of portfolios were divided. Feedback from school principals ranged from the compilation of portfolios as

extra work and of little benefit to a helpful process to improve performance. Different school principals perceived their work on portfolios either as a compliance tool, documenting their past achievements, or as a forward-looking tool for goal-setting and professional development. Johnston and Thomas argued that it was the contextualisation of portfolios in a larger supportive social network of professional practice that enabled portfolios to function as a learning tool. They, ultimately, suggested that, considering the time and efforts involved, portfolios can mainly serve as a tool for formative appraisal.

Testing

Problem-based tests, i.e. the simulation of particular leadership situations in case studies that test the responses of school leaders, constitute a further way of evaluating leadership capacity through observation. While some research has been conducted on the development of leadership tests (e.g. Claudet, 2002), in practice very few systems and none of the countries that provided information as part of the OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes make use of test settings to inform appraisal (see Table 7.A.1).

The limited research on leadership tests suggests that this tool may primarily be useful to inform formative appraisal. Tests may allow evaluators to assess a leaders' ability to respond to a specific challenge, but they may not necessarily relate to their everyday practice (Pashiardis and Brauckman, 2008). These concerns about the accurate assessment of a school leader's performance may limit the value of leadership tests for appraisal. However, tests may add value to appraisal when used in combination with other instruments, especially to inform professional development (Claudet, 2002).

Stakeholder surveys and questionnaires

School leader appraisal may also take the views of stakeholders (e.g. teachers, parents and students) into account. The use of stakeholder surveys and questionnaires may serve to provide an additional perspective on a school leader's performance and to provide information on the level of stakeholder satisfaction. Depending on the purposes and uses of an appraisal system, the results of surveys and questionnaires may be shared with the school leader, evaluators, and/or the whole school community.

While appraisal in a few countries may include the views of stakeholders (e.g. various states and territories in Australia, various provinces and territories in Canada, Chile [Performance Appraisal], Korea [School Principal Appraisal for Professional Development and Evaluation for School Management], New Zealand and Poland), most countries do not (see Table 7.A.1). Some states and territories in Australia draw on information collected through parent surveys for the appraisal of school principals and on staff surveys when appraising deputy school principals. The province-wide appraisal framework in Ontario, Canada, leaves it to the discretion of school leaders to include stakeholder views of staff, parents and/or students depending on a school leader's objectives. For example, if a school principal's Performance Plan includes the objective of increasing parent involvement in student learning, a survey of parents before and after the implementation of strategies for this goal could be used to measure relevant outcomes. The Korean School Principal Appraisal for Professional Development and Evaluation for School Management systems draw heavily on information gathered through teacher, student and parent surveys. School boards of trustees in New Zealand may also use teacher, parent and/or student surveys to form a judgement of school principals'

performance. In Poland, evaluators are required to take the judgement of the school board or the pedagogical council of the school as well as trade unions into account.

Much of the literature on school leader appraisal identifies wider stakeholder involvement as a good practice to increase the effectiveness of appraisal by assessing performance from a comprehensive perspective (Reeves, 2009; Leon et al., 2011). Including stakeholders' views as one source of information in the appraisal of school leaders also reflects school leaders' tasks and responsibilities that typically include engagement with the wider community. Stakeholders in some contexts have, therefore, supported arguments to have a voice in the appraisal of school leaders. In Belgium (Flemish Community), for instance, the Flemish Student Council (*Vlaamse Scholieren Koepel* [*VSK*]), advocated for a greater student voice in the appraisal of Flemish school leaders, e.g. through local student councils (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

However, the use of stakeholder surveys and questionnaires also entails practical challenges. When using surveys and/or questionnaires, evaluators forming an ultimate judgement of a school leader's performance need to bear issues of fairness and objectivity, validity and reliability in mind, especially where appraisal may lead to summative consequences. Different stakeholders may hold different expectations of their school leaders depending on their interests, levels of experience and involvement in the school's affairs (Ginsberg and Thompson, 1992; Davis and Hensley, 1999). Not all decisions a school leader makes may be viewed equally favourable by all stakeholders (Reeves, 2009). Owing to their level of involvement in the everyday operations of schools and direct experience of school leaders' practices and behaviours, teachers may provide the most valid information on a school leader's performance (Marzano et al., 2005).

Student outcomes

School leader appraisal may also rely on student achievement data. This may include various types of student-related information, such as the percentage of students achieving a certain mark, the percentage of a particular group of students achieving a certain mark (e.g. students at risk), the aggregate mean of student marks across one school, improvement of aggregate means of student marks or a comparison of the aggregate mean of student marks across similar schools, student dropout and repetition rates (Duke, 1992).

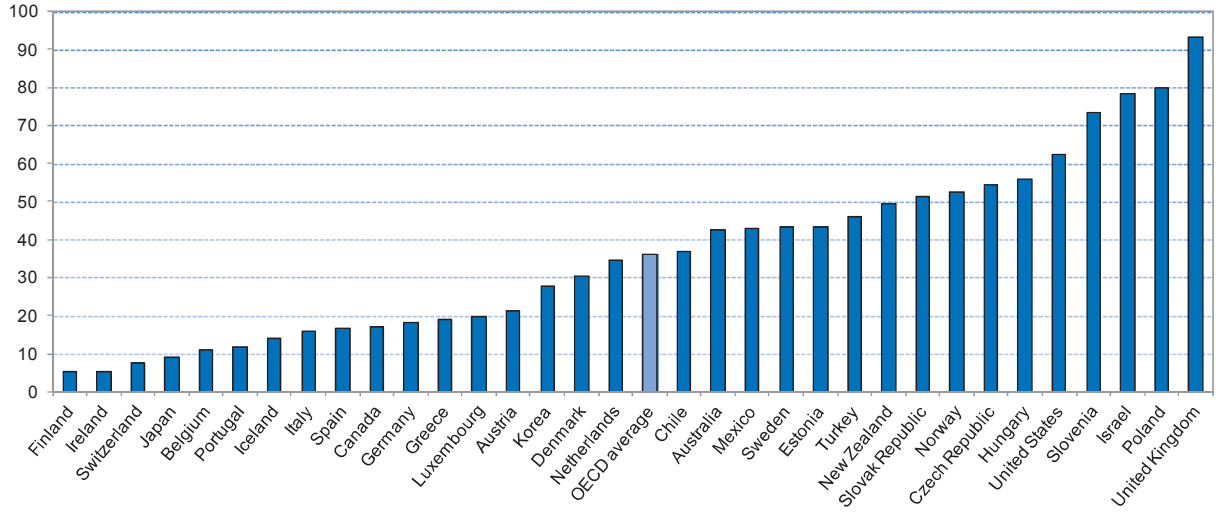
There are, however, only a few systems that require or give scope for evaluators to take student outcomes into consideration. These include states and territories in Australia, various provinces and territories in Canada, Chile (Performance Appraisal), France (ISCED levels 2 and 3), Korea (Evaluation for School Management), and New Zealand. To provide a few specific examples, in Ontario, Canada, school leaders may include student achievement data as one source of information as well as qualitative data provided by teachers, such as student writing samples. Mexico (centrally managed schools at ISCED level 3) constitutes the only country where the appraisal of school principals is mostly based on student outcomes.

Data collected through PISA provide an insight into the extent to which school principals themselves reported that student achievement data was used in their appraisal (see Figure 7.6). It is, however, important to bear in mind that this information rests on school principals' interpretation of the questionnaire and that school principals' may have included informal appraisal and feedback practices at a local level in their responses. Across OECD countries, on average only 36% of students went to schools where school principals reported that student achievement data was used in the appraisal of their

performance (OECD, 2010). Countries in which students went to schools with a larger percentage of school principals reporting the use of achievement data in their appraisal included the United Kingdom (93%), Poland (80%), Israel (78%),[30] Slovenia (74%) and the United States (63%). Canada (17%), Spain (17%), Portugal (12%) and Belgium (11%) constituted examples where a smaller percentage of students attended schools in which school principals reported the use of student achievement data in their appraisal.

Figure 7.6 Use of student achievement data in the appraisal of school principals' performance (2009)

Percentage of students in schools whose school principals reported that student achievement data was used in the appraisal of their performance (%)



StatLink http://dx.doi.org/10.1787/888932791571

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris.

Research on the effects of appraisal that takes student outcomes into account on school leaders' practices and behaviours or learning outcomes is limited. Arguments on the use of student outcomes as a source of information for appraisal often rest on research about the impact of effective school leadership on student achievement. Given school leaders' importance for creating effective teaching and learning environments, there may be grounds for taking school leaders' efforts to contribute to better outcomes for all students into account (Brown-Sims, 2010). Parylo et al. (2012b), furthermore, suggested that consideration of student achievement data may help make appraisal more independent of an evaluator's judgement that inevitably involves a subjective dimension. Based on such arguments, appraisal models developed by the non-profit organisation New Leaders (New Leaders for New Schools, 2010) and the Wallace Foundation (Murphy et al., 2006; Porter et al., 2006; Goldring et al., 2007) in the United States advocate holding school leaders in part directly accountable for the performance of students in their school. New Leaders for New Schools (2010), for example, suggested basing 70% of the appraisal judgement on student outcomes. In the United States, federal funding initiatives in education linked to the American Reinvestment and Recovery Act (2009) also often require states to tie appraisal results for both teachers and school leaders to student achievement (e.g. Race to the Top initiative). Chicago, Seattle and Delaware

constitute some of the systems in the United States that take student test scores into account (Portin et al., 2006).

However, other researchers have raised caution about an overemphasis on student outcomes as a source of information for appraisal considering school leaders' mainly indirect effects on student outcomes and the range of factors that impact student achievement over which school leaders may have little or no control (Duke, 1992; Normore, 2004; Pashiardis and Brauckman, 2008; Reeves, 2009). Pont et al. (2008a) stressed that student learning is influenced by many factors including school organisation, resources and climate and curriculum structure and content, but also teacher skills, knowledge, attitudes and practices; students' skills, expectations, motivation and behaviour; peer group skills, attitudes and behaviour; and family resources, attitudes and support. Spillane and Kenney (2012) have also recently pointed out that school leaders and teachers stand in an interdependent relationship in which school leaders need to engage in teachers' willingness to change to achieve an improvement in teaching and learning. Considering the role of school leaders for developing, supporting and supervising teachers as part of pedagogical leadership, research on the potentially detrimental effects of test-based accountability mechanisms for teachers equally points towards some of the risks involved (see Chapter 5 for an in-depth discussion of the use of student outcomes for teacher appraisal). Considering the risks and limitations involved, Reeves (2009) suggested complementing student achievement data with additional information such as surveys, questionnaires and/or the observation of a school leader's focus on student success, quality instruction and a rigorous curriculum. He also pointed out that, considering that temporarily poor student results may hide excellent leadership, evaluators should take into account a school leader's actions and steps following the publication of students' marks.

Given school leaders' role for improving teaching and learning for all students, but the wide range of factors within and outside schools that impact on students' learning and the largely indirect effects of effective school leadership, policy makers face the challenge of findings ways to account for student outcomes within an appraisal framework that respect these limitations and ensure an accurate, fair, valid and reliable appraisal process.

Capacity

This section analyses countries' approaches to the distribution of responsibilities for the implementation of appraisal. It also includes a discussion of the necessary competences and capacity of both evaluators and school leaders to perform an appraisal and to use results for the improvement of school leaders' practices and behaviours.

Distributing responsibilities for the implementation of the appraisal process

Embedded within the overall allocation of responsibilities (see section on the governance of appraisal schemes), evaluators from central/state education authorities, intermediate agencies, such as school inspectorates, regional education authorities, local education authorities, school organising bodies or school boards carry responsibility for implementing the appraisal process in an individual capacity or as part of a team of evaluators. The appraisal process may additionally involve external consultants and/or peer school leaders from another school. Responsibilities for evaluating school leaders and the composition of an evaluator team may differ depending on the school leadership position.

In a first group of countries, the evaluators of school principals come from central/state education authorities (e.g. various states and territories in Australia, Belgium [French Community] [public schools], various provinces and territories in Canada and Mexico [centrally managed schools at ISCED level 3]), from intermediate agencies (e.g. Israel [ISCED level 1 and partially ISCED level 2] and various autonomous communities in Spain) and/or from regional authorities (e.g. France [ISCED levels 2 and 3], Korea [Evaluation for School Management and School Principal Appraisal for Professional Development] and Portugal). In a second group of countries, the evaluators of school principals belong to school organising bodies, school boards and/or local education authorities (e.g. various states and territories in Australia, Belgium [Flemish Community], Belgium [French Community] [government-dependent private schools], various provinces and territories in Canada, Chile [Performance Appraisal], New Zealand, Slovenia, and Northern Ireland in the United Kingdom). In Poland, local and/or regional education authorities are responsible for appraising school principals. In France (ISCED level 1), evaluators belong to local inspection bodies (IEN) and regional education authorities.

Depending on the context, an individual evaluator (e.g. Belgium [Flemish Community]) or a team of evaluators (e.g. Korea [School Principal Appraisal for Professional Development], New Zealand, Portugal and Northern Ireland in the United Kingdom) are responsible for carrying out the appraisal. In New Zealand, the good practice framework for school principal appraisal recommends the involvement of two different evaluators to judge a school principal's performance in a summative fashion and to provide formative feedback and support respectively. In some systems, evaluators or evaluator teams are free to or required to draw on the additional expertise of external consultants (e.g. Belgium [French Community] [government-dependent private schools], New Zealand, and Northern Ireland in the United Kingdom) or external peer school leaders (e.g. Korea [School Principal Appraisal for Professional Development]). In Northern Ireland in the United Kingdom, for example, school principals are appraised by two reviewers from the Board of Governors. This team of evaluators is assisted by an External Advisor. External Advisers are nominated by the employing authority of the school and trained and accredited by the Regional Training Unit.[31] They are responsible for providing support to governor reviewers when setting performance objectives, when evaluating a principal's performance at the end of the appraisal cycle, and when identifying personal or professional training needs (Department of Education, Northern Ireland, forthcoming).

Concerning the appraisal of deputy school principals, evaluators are at times equivalent to those evaluating school principals (e.g. Belgium [French Community], France [ISCED levels 1, 2 and 3] and Northern Ireland in the United Kingdom). In other contexts, school principals carry responsibility for appraising their deputy school leaders (e.g. various states and territories in Australia, including Victoria, various provinces and territories in Canada, Chile [Performance Appraisal], and Israel [ISCED level 1 and partially ISCED level 2]).

With regards to the locus of authority for implementing an appraisal process, reliance on local evaluators holds both benefits and drawbacks. On the one hand, evaluators from local education authorities often bring in-depth knowledge of the particular circumstances of a certain school and a school leader's work, something particularly important considering the role of context for successful school leadership. Thanks to evaluators' insights into local contexts, scope for the local implementation of appraisal also constitutes an opportunity to create trust, commitment and ownership. On the other hand,

local agency for implementing an appraisal may limit the capacity for the system-wide implementation of an accurate, fair, valid and reliable appraisal process for all school leaders. As the professional organisation for school principals in New Zealand pointed out the quality of school leader appraisal in New Zealand differs greatly depending on the capacity of the responsible school board of trustees. In some cases, members of the school board appraising a school leader may lack preparation and the background and knowledge to provide professional advice to school leaders (Nusche et al., 2012). In Belgium (Flemish Community) similar concerns have been raised. Evaluators that belong to school boards of government-dependent private schools that mostly comprise volunteers may lack the professional knowledge and skills to appraise school leaders effectively (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010). These considerations, then, highlight the benefits of involving local expertise, but raise the challenge of ensuring the development of the necessary skills and competences at a local level to successfully implement an appraisal process.

Even though the relative costs and benefits in terms of time, resources and added value to the appraisal process need to be taken into account, research has also pointed out the importance of involving multiple evaluators to draw on different perspectives on a school leader's performance and to increase the accuracy, fairness, reliability and validity of an appraisal (Reeves, 2009; Parylo et al., 2012b). External consultants/advisors and peer school leaders may both contribute additional expertise to an appraisal process:

- As some research indicates, external consultants or advisers can add additional expertise and help create good relationships between appraisers and appraisees. Crawford and Earley (2004) examined the views of external advisors on their own role in the headteacher appraisal process in England. As Crawford and Earley suggested, external advisors, often bringing extensive educational experience and expertise, may be helpful for school boards and headteachers to develop a productive relationship. They may also help to encourage headteachers to discuss issues they would not normally share with others. Their limited contact time with schools, however, may at the same time limit their potentially positive impact on the appraisal process.
- External peer school leaders may equally provide a further viewpoint for evaluators. The limited evidence on the effects of peer appraisal suggests the benefits of this approach as an opportunity for sharing good practices and for learning from each other, thus strengthening the formative dimension of an appraisal. However, issues of legitimacy, distance, and trust need to be addressed, especially in the case of appraisal procedures that may lead to summative consequences (Pashiardis and Brauckman, 2008; Parylo et al., 2012b). Research on the experience of headteacher appraisal in England in the 1990s that involved peer school leaders found that headteachers viewed the involvement of peer school leaders in an overall positive light (Cullen, 1997; Hellawell and Hancock, 1998). In both studies headteachers particularly valued their insights into leadership practices and behaviours when themselves acting as peer evaluators of other headteachers. Involving a peer school leader in the appraisal process also corresponds to the development of school leadership models in which school leaders assume responsibilities beyond their own school and engage in the improvement of the wider education system (e.g. through school partnerships and networks and as critical friends).

Developing evaluators' and school leaders' competencies for appraisal

A further aspect of appraisal relates to the required competencies of both evaluators and school leaders to carry out an appraisal process and to make best use of its results as well as the ways in which the required competences are developed most effectively.

In many contexts, policy makers have strengthened initiatives to support school leaders' professional development (also see section on Impact, drivers and contextual development for comprehensive approaches to school leadership development; also see, for example, Halász (2009) for research analysing school leaders' professional development in five central European countries). To provide two examples, Ireland has introduced a Leadership Development for Schools (LDS) programme that provides professional development opportunities with a focus on improving teaching and learning for all students for emerging and established school leaders on an annual basis through a variety of formats (e.g. interactive seminars, active learning networks and virtual learning environment).[32] Norway has implemented a National Training Programme for Principals in 2009 that aims to support school principals to become pedagogical leaders that work towards improving teaching in learning in schools. Initially targeted at new school principals, 621 participants took part in the programme that can be taken part-time over one-and-a-half to two years between 2009 and 2011 (Norwegian Directorate for Education and Training, 2011).

However, throughout OECD countries little is known about the ways in which evaluators and school leaders receive appraisal-related training or about the extent to which professional development for school leaders includes appraisal-related competencies. The Northern Territory in Australia has introduced a coaching model that involves evaluators and school principals as part of its School Accountability and Performance Improvement Framework. This coaching model aims to encourage school principals' commitment for the wider school evaluation, school self-evaluation and individual appraisal process and to develop related competencies. The province-wide appraisal system in Ontario, Canada, requires district school boards to provide training for evaluators to ensure effective implementation. Appraisal-related training is also offered through school leader and supervisory officer associations as well as through web casts. Northern Ireland in the United Kingdom requires External Advisers to apply for annual reaccreditation to ensure they are informed about potential changes to the Performance Review and Staff Development Scheme and that they have the knowledge and competencies to provide helpful advice to governor reviewers. As very limited research, however, indicates, policy making in some contexts lacks attention towards the development of appraisal-related knowledge and competencies among evaluators and school leaders. According to the OECD Improving School Leadership project, for example, school boards that carry responsibilities for appraisal in many countries, often lack support structures and opportunities to develop related competencies (Pont et al., 2008a). A number of school principals that took part in Parylo et al.'s (2012b) study also criticised a lack of professional development and support to prepare them and their appraisers for the implementation of appraisal.

Research has stressed the role of evaluators' and school leaders' appraisal-related competencies for the consistent and effective implementation of an appraisal system with fair, valid and reliable procedures and results (Kimball et al., 2009). Competent and credible evaluators also help build trust among school leaders in the appraisal system. All of those involved in appraisal, including evaluators, external advisors, peer school leaders and school leaders themselves, therefore, require adequate information about an appraisal

process and related training and professional development opportunities (Kimball et al., 2009; Clifford and Ross, 2011; Leon et al., 2011; Sanders and Kearney, 2011). Considering the key role of capacity for effective appraisal, policy makers face the challenge of providing opportunities for capacity development and of ensuring an adequate investment of resources and time in the development of evaluators' and school leaders' competencies for appraisal.

However, there is little research or guidance available on evaluators' and school leaders' required competencies and how these are best developed (Clifford and Ross, 2011). Piggot-Irvine (2003) provided some insights into the competencies evaluators need to successfully implement an appraisal process. As she pointed out, evaluators need to be knowledgeable about the values and purposes of an appraisal scheme. They also need to be familiar with the instruments that are used (e.g. how to set objectives, how to gather data, how to interview and observe school leaders). Appraisal frameworks that involve the selection of sources of information, in particular, require capacity to select an effective, valid and reliable mix of instruments. Further necessary competencies include the ability to ensure a confidential approach to information, but also to ensure transparency about the way the appraisal judgement was formed. According to Piggot-Irvine, evaluators need to be trained in approaches to establish respectful, trust-based and open relationships with school leaders. The contextual nature of effective school leadership requires evaluators to be sensitive to contexts and a school leader's experience when forming a judgement about a school leader's performance (Goldring et al., 2007; Pashiardis and Brauckman, 2008). Training for school leaders needs to ensure that they are knowledgeable about the appraisal process and the expectations that they are appraised against (Ginsberg and Thompson, 1992; Clifford and Ross, 2011). For the developmental function of appraisal, school leaders need to be competent to use and interpret appraisal results for their own improvement.

Use of results

Depending on the purposes of an appraisal system, countries use appraisal results to inform formative and summative decisions. Results may provide developmental feedback for school leaders, be linked to their professional development, to employment-related decisions (e.g. contract renewal, reassignment and dismissal) and to school leaders' career advancement (e.g. on a multilevel career ladder). Appraisal results may also provide a basis for allocating financial rewards to create incentives for high performance (e.g. through a one-off financial bonus). However, it is essential to bear in mind that summative decisions are always embedded within the boundaries of specific contractual and employment frameworks, depend on the distribution of responsibilities for employing school leaders and are often subject to agreements between social partners (see Table 7.A.2 for an overview of the employment status of school leaders). The systematic collection and analysis of appraisal results may, furthermore, provide information at a system-level for the further development of policies to strengthen leadership in schools.

While some countries appraise their school leaders for solely formative or solely summative purposes, the majority of countries combine both developmental and accountability functions (see Table 7.2 and Table 7.A.2). An increasing number of countries conducts appraisal to provide feedback and to inform professional development. In many of these cases, appraisal may lead to summative consequences for underperforming school leaders if a school leader fails to improve after an unsatisfactory rating and the introduction of an improvement plan, for example. Opportunities for the career development of school

leaders are still scarce and only in a few countries can school leaders progress on a multilevel career structure based on considerations that include appraisal results (see Table 7.A.2). The ways in which results are used may differ depending on a school leader's performance in the appraisal as well as the school level and school type.

Table 7.2 The use of school leader appraisal results

Professional development	Career advancement				
		Does not inform career advancement	Expected to or may influence/inform career development	Linked to salary progression within single salary scale or in a multilevel career ladder	Linked to career advancement in a multilevel career ladder
	Does not inform professional development		Korea (Evaluation for School Management)	France (ISCED levels 1, 2 and 3), Portugal, various autonomous communities in Spain	France (ISCED levels 1, 2 and 3),
	Expected to or may influence professional development	Belgium (French Community)	Belgium (Flemish Community)	New Zealand, Northern Ireland (United Kingdom)	Various provinces and territories in Canada, Israel (ISCED level 1 and partially ISCED level 2)
	Systematically linked with professional development	Chile (Performance Appraisal), Korea (School Principal Appraisal for Professional Development)		Mexico (centrally managed schools at ISCED level 3)	Various states and territories in Australia, various provinces and territories in Canada

Note: This taxonomy only includes countries which submitted information based on a questionnaire developed by the OECD Review of Evaluation and Assessment Frameworks for Improving School Outcomes. It does not consider the use of appraisal results for contract renewal or rewards purposes. Countries where the design of appraisal procedures is at the discretion of local education authorities, school boards or school organising bodies are not included. Countries with various requirements for the appraisal of school principals and countries with multiple jurisdictions have been included in all respective cells to reflect differences across sub-national jurisdictions.

Using appraisal results for formative purposes

Appraisal itself can be used as a formative process that provides feedback for school leaders about their strengths and weaknesses. Through the identification of areas of improvement, appraisal results can, furthermore, feed into the preparation of individual development plans and/or inform the professional development opportunities a school leader takes part in.

Professional development

While in a few systems appraisal is systematically linked to professional development plans and activities (e.g. various Australian states and territories, various provinces and territories in Canada, Chile [Performance Appraisal], Korea [School Principal Appraisal for Professional Development] and Mexico [centrally managed schools at ISCED level 3]) and in various contexts appraisal is expected to influence opportunities for professional development (e.g. Belgium [Flemish and French Communities], various provinces and territories in Canada, Israel [ISCED level 1 and partially ISCED level 2], New Zealand, and Northern Ireland in the United Kingdom), in various other systems appraisal results do not have any links to professional development programmes (e.g. France [ISCED levels 1, 2 and 3], Poland, Portugal, Slovenia and various autonomous communities in Spain). The recently developed school leader appraisal models in Chile (Performance Appraisal); Ontario, Canada; and Victoria, Australia, constitute examples of appraisal schemes that are explicitly designed to improve school

leaders' practices and behaviours through formative feedback and professional development. In all of these three appraisal frameworks, appraisal is based on and systematically tied to professional development plans that are embedded within wider school goals. In Ontario, Canada, and Victoria, Australia, school leader appraisal is also firmly grounded within a wider system of professional development for school leaders (see Box 7.3). Four countries participating in the OECD Review reported a further interesting way of strengthening a formative dimension in the appraisal of their school leaders. Chile (Performance Appraisal), New Zealand, Portugal and Northern Ireland in the United Kingdom require the appraisal of a school leader's engagement and participation in professional development activities as one of the appraisal aspects and criteria.

In some systems, additional feedback and/or professional development opportunities may be offered to school leaders depending on their appraisal score, i.e. the rating that is awarded or the performance threshold that is reached. High performers may be rewarded with additional opportunities for professional development (e.g. various Australian states and territories, and Korea [School Principal Appraisal for Professional Development]). School leaders with an unsatisfactory performance may also be required or advised to complete compulsory training to improve (e.g. various Australian states and territories, various provinces and territories in Canada, Chile [Performance Appraisal], Korea [School Principal Appraisal for Professional Development], New Zealand, Portugal, and Northern Ireland in the United Kingdom).

However, the extent to which appraisal results inform professional development in a useful and meaningful way in practice may differ from formal requirements. This also depends on evaluators' and school leaders' capacity to use appraisal results. A review of principal appraisal in school districts in the United States, for example, found that almost half of the appraisal models reviewed failed to provide school leaders with clear feedback that was linked to a development plan on what they could be doing better to improve teaching and learning in their school (Goldring et al., 2008). Even if school leader appraisal results successfully feed into a school leader's professional development, opportunities for professional development may fail to support school leaders in the improvement of their leadership practices and behaviours and the development of their pedagogical leadership. School leaders in various OECD countries have often reported a feeling of inadequate preparation for their roles, a lack of continuous feedback on their performance, and a lack of opportunities for professional development (Pont et al., 2008a). To give an example, the management of resources needs to be sufficiently aligned with pedagogical purposes to improve teaching and learning. However, school leaders in many contexts often lack adequate training opportunities to enable them to do so (Pont et al., 2008a).

Using appraisal results for summative purposes

School leader appraisal results can also be used as the basis for employment-related decisions, to inform a school leader's salary progression and/or career advancement, and the allocation of financial rewards for successful school leaders.

Employment-related decisions

Concerning the use of appraisal results for contract renewal processes, one can distinguish between two groups of countries. In a first group of countries (e.g. Czech Republic, Poland, and Slovak Republic) appraisal is exclusively linked to inform

employment-related decisions and the contract renewal process in case a school leader decides to reapply for his or her position. The Czech Republic, for instance, passed new legislation in 2012 that changed the regulations for the dismissal of head teachers. While school organising bodies could previously only dismiss head teachers in the case of gross violations or their failure to fulfil their legal obligations, school organising bodies now decide on head teachers' reappointment after a term of six years, as informed by the appraisal process (IIE [Institute for Information on Education], 2011). In a second group of countries, the results of appraisal processes that combine both formative and summative functions may also be used to inform the contract renewal process, for instance in the case of school leaders employed on a fixed-term contract (e.g. some states and territories in Australia, including Victoria, and Mexico [centrally managed schools at ISCED level 3]). In Mexico, for example, the appraisal system may inform whether school principals are granted a permanent contract or not.

In a number of countries, including those where appraisal primarily serves formative ends, appraisal may lead to employment-related consequences for school leaders that are judged as performing below the expected standards (e.g. dismissal and transfer to another school or leadership position). Two countries indicated that underperformance can result in a school leader's transfer (various states and territories in Australia and various autonomous communities in Spain). Various other countries indicated that underperformance entails the possibility to dismiss a school leader or to initiate a separate procedure that may lead to a school leader's dismissal in the case of his or her continued underperformance. This includes various states and territories in Australia, Belgium (Flemish Community), various provinces and territories in Canada, Chile (Performance Appraisal), Israel (ISCED level 1 and partially ISCED level 2), New Zealand, and Northern Ireland in the United Kingdom. In the Flemish Community of Belgium, for example, school principals on temporary contracts may be dismissed if their performance is rated as unsatisfactory. They may, however, remain employed in a teaching capacity if they had previously been granted a permanent teaching contract (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

Appraisal often involves various steps before it leads to a school leader's dismissal or results in a separate dismissal process. In Ontario, Canada, for example, school principals and vice-principals can be re-assigned to a different leadership position or dismissed from their responsibilities if appraisal has resulted in a third unsatisfactory rating and after the introduction of improvement plans. In New Zealand, school principals are typically employed on a permanent basis. In case of serious concerns, however, a competency process designed to provide remedial support may be initiated. If concerns about a school principal's performance persist, the competency process may result in a school principal's dismissal.

Career advancement

Very few countries have established opportunities for school leaders' career advancement (see Table 7.A.2). While various countries have no opportunities for career advancement in place, their school leaders can receive salary increments within a single salary scale depending on factors such as their length of service and the size of their school (e.g. Austria, Belgium [Flemish and French Communities], Chile, Czech Republic, Finland, Iceland, Ireland, Italy, Luxembourg, Mexico [ISCED level 3], Netherlands, New Zealand, Poland, Portugal, various autonomous communities in Spain, and Northern Ireland in the United Kingdom). Five countries with central/state frameworks for appraisal reported that appraisal results may be taken into account as one factor when

deciding about a school leader's progression in the salary scale (France [ISCED levels 1, 2 and 3], Mexico [centrally managed public schools at ISCED level 3], New Zealand, Portugal, and Northern Ireland in the United Kingdom). In New Zealand, for example, appraisal results may influence a school leader's advancement on the single salary scale together with the length of service. In Northern Ireland in the United Kingdom, the results of the Performance Review and Staff Development Scheme gathered in the review statement are part of the body of evidence used to inform pay progression decisions. While in various autonomous communities in Spain the salary scale is only based on one salary step, appraisal results may be used to decide about the consolidation of this salary level. Among countries where school leader appraisal is at the discretion of local education authorities, school organising bodies or schools boards, the Netherlands indicated that appraisal results may be taken into account in a school leader's salary progression. Typically, however, the remuneration of school leaders with similar qualifications working at the same school level does not consider school leader's particular working conditions and their performance and commitment (Pont et al., 2008a). As Eurydice pointed out with regards to school heads' statutory salaries in European countries, the overall potential for increases in school heads' salaries is often not exceptional and school head's statutory salaries are often influenced by the enrolment at schools (EACEA, 2012).

Only a very small number of countries give school leaders the opportunity to develop on a multilevel career ladder within schools that are linked to a single or a separate salary scale (e.g. various states and territories in Australia, various provinces and territories in Canada, France [ISCED levels 1, 2 and 3], Hungary, Israel, Korea, Mexico [ISCED levels 1 and 2], and Norway). Only in a few of these countries career advancement is linked to the results of appraisal that is governed through a central/state policy framework, often together with several other factors such as a school leader's length of service and completion of professional development (e.g. various states and territories in Australia, various provinces and territories in Canada, France [ISCED levels 1, 2 and 3], and Israel [ISCED levels 1 and 2]). In all of these countries except Israel, appraisal may lead to the deferral of a school leader's promotion in case his or her performance is judged as unsatisfactory. In Korea (Evaluation for School Management), a school leader's promotion on the multilevel career structure may also take appraisal results into account. Among countries where school leader appraisal is entirely at the discretion of local education authorities, school organising bodies or schools boards, or where appraisal procedures are defined at a local level, Hungary (e.g. at re-appointment as school principal)[33] and Norway indicated that school leader appraisal results may be taken into account in a school leader's career advancement.

Performance rewards and sanctions

Appraisal may provide a basis for creating financial incentives, such as differentiated pay or a financial bonus for outstanding performance, and sanctions in the case of underperformance. Only very few countries, however, link appraisal results to financial rewards and/or sanctions. In various states and territories in Australia and France (ISCED levels 1, 2 and 3) school principals and deputy principals may receive a permanent salary increment as a reward for outstanding performance. In France (ISCED levels 1, 2 and 3), the exceptional performance of school principals and deputy school principals may also be rewarded with a one-off financial bonus. In Chile (Performance Appraisal), outstanding school leaders may receive a salary increment for a fixed period of time. In New Zealand, exceptional school leaders may be rewarded with a staged career allowance. In Northern

Ireland in the United Kingdom, Boards of Governors may reward school principals that demonstrate a very high performance with an additional point pay award in the salary scale.

The School Management process in Korea and the school principal appraisal system in Slovenia represent central appraisal frameworks that specifically aim to provide financial incentives and rewards to encourage effective school leadership. In Korea, the Evaluation for School Management provides the basis for rewarding outstanding school leaders with performance-based rewards and for financially sanctioning school leaders that score among the bottom 2 to 3% (Kim et al., 2010). In Slovenia, appraisal is used to reward successful school principals with a one-off financial bonus that can reach up to two basic monthly salaries. The amount of the financial bonus is determined by the respective school council when approving the school's annual report for the previous year (Brejc et al., 2011).

The effective use of appraisal results is essential for making appraisal a meaningful exercise

As some educational researchers have stressed, ensuring the effective use of appraisal results constitutes a key element of appraisal to make it a meaningful process (Reeves, 2009). Limited research provides some insights into the ways in which appraisal results can be used more effectively and the benefits and drawbacks of different approaches to the potential formative and summative uses of results.

When considering the ways in which appraisal results are used, the current challenges facing the school leadership profession (see section on impact, drivers and contextual developments) and the contextual dimension of effective school leadership (see section on the analytical approach of this chapter) need to be taken into account. The Wallace Foundation (2009) stressed the opportunity of tailoring appraisal to local contexts and individual needs when using appraisal results for formative and/or summative decisions as well as the importance of taking local contexts and individual needs into account in the process. For formative uses, evaluators need to consider school leaders' specific strengths and weaknesses as well as school contexts to tailor professional development plans to these circumstances. Pashiardis and Brauckman (2008) argued that due consideration to a school leader's experience can significantly improve the usefulness of the appraisal process since the needs of school leaders change markedly as they progress through their career. Pashiardis and Brauckman, therefore, suggested adjusting the feedback provisions for school leaders to the stages of pre-leadership, initial, experienced and exiting leadership. For summative uses, evaluators equally need to take into account specific school and individual contexts and a school leader's experience to ensure a fair appraisal process.

Concerning the formative use of results, some research has pointed out that appraisal as an opportunity for feedback as well as the use of appraisal results for professional development may be particularly valued by school leaders. A few studies that have examined school leaders' views of appraisal provide some insights into the value of using appraisal itself as a tool for feedback and to inform professional development. Parylo et al.'s (2012b) research on school principals' experience of appraisal highlighted school leaders' appreciation of honest and constructive feedback and an open dialogue about their strengths and weaknesses throughout the appraisal process to improve their practices and behaviours. Considering the position of school leaders within their school, school leaders may, furthermore, lack opportunities to receive professional feedback from other, more informal, sources, such as peers and face feelings of professional isolation. As Hellawell and Hancock's (1998) and Clifford and Ross (2011) pointed out, appraisal that

provides meaningful feedback may constitute one strategy for tackling the risk of school leaders' sense of professional isolation. Headteachers in Hellawell and Hancock's study on the experience of appraisal in England valued the process to overcome professional isolation as well as an opportunity for critical reflection and encouragement. Conversely, a lack of feedback from the appraisal process or a failure to effectively feed into school leaders' professional development may give school leaders the feeling of a lack of care and interest in their work (Reeves, 2009).

Further research has pointed out that, depending on the way appraisal is designed and implemented, school leaders may perceive appraisal as a meaningless exercise that did not help them develop professionally (e.g. Davis and Hensley, 1999; Gaziel, 2008). These studies indicated that school leaders valued appraisal with a strong formative dimension as an opportunity to receive an outside perspective on their work. Sun and Youngs (2009) concluded from their research that, in order to effectively influence school leader behaviours and practices, appraisal needs to go beyond employment-related decisions (e.g. contract renewal or salary allocations) and, instead, promote professional development, accountability and school restructuring. Considering countries' policy frameworks and practices on the ground, policy makers need to find ways to establish feedback as a core element of appraisal itself, channels for appraisal to effectively feed into professional development and meaningful professional development opportunities to ultimately improve school leadership practices and behaviours.

Concerning the summative use of results, there is little empirical evidence on the specific effects of the use of appraisal for career advancement, for determining school leaders' salaries or for allocating financial bonuses to outstanding school leaders on school leaders' practices and behaviours as well as schools as a whole. Considering the difficulty of attracting qualified candidates and retaining successful school leaders over time in many contexts, some research suggests that appraisal may constitute one policy lever for creating incentives to facilitate the recruitment and retainment of effective school leaders (Pont et al., 2008a; Reeves, 2009). Pont et al. (2008a) suggested that rewards for high performance can be an effective tool for raising motivations when understood as a form of positive feedback. Some school districts in the United States have established appraisal procedures that include financial rewards as an integral part of the use of appraisal results. In the school district of Pittsburgh, Pennsylvania, for example, the Pittsburgh Urban Leadership System for Excellence entails the use of appraisal results to determine performance-based pay.[34] With the introduction of this system, compensation is no longer based on annual salary increases for all school principals or a system based on salary steps determined by length of service. Under the new system, school principals can earn an annual salary increment of up to USD 2 000 that becomes part of the base salary if they are rated proficient across the seven performance standards and 27 components of practice laid out in the district's performance standards or if rated as satisfactory in completing the professional growth project. School principals can earn an additional achievement bonus of up to USD 10 000 for raising student achievement that is not retained in the base salary. The effects of such schemes will, however, vary among individuals and contexts, and also depend on the career stage a leader is in. Links between appraisal and school leaders' salaries and financial bonuses, in particular, are contentious. Research on performance-based pay for teachers is difficult and has produced mixed results, pointing to similar difficulties for the case of school leaders (see Chapter 5). Using appraisal for salary decisions may also have negative effects on the school climate, teamwork and collaboration among staff and a school leader's internal motivation (Pashiardis and Brauckman, 2008; Pont et al., 2008a).

While appraisal may be used to determine school leaders' career advancement, to inform salary decisions and/or to award performance-related financial rewards to recognise and reward outstanding school leadership, policy makers need to keep in mind that the evidence base for such schemes is rather mixed and that strong ties between appraisal and rewards are controversial. If appraisal informs pay decisions, a sound appraisal system must be in place, which includes clear performance criteria, reliable indicators, competent and trained evaluators and due attention to the particular contexts in which school leaders work (Pont et al., 2008a). In England, headteachers' salaries are largely related to the number of students of a school with some flexibility to link pay to performance through an incremental pay scale. A study by PriceWaterhouseCoopers (2007) on school leadership in England suggested that 70% of primary headteachers and 81% of secondary headteachers agreed that individual performance should be taken into account in determining pay, but raised concerns about the effectiveness of the current system in place. For systems that use appraisal to inform summative decisions, one challenge, therefore, lies in establishing fair, valid and reliable appraisal procedures that can provide an adequate basis for such decisions. However, as Glasman and Glasman (2010) pointed out it is also essential to provide school leaders with opportunities and support to improve their performance before taking any summative decision (e.g. contract renewal and dismissal).

Pointers for future policy development

This chapter has reviewed the approaches countries are taking to appraising the performance of their individual school leaders in light of available research and evidence. The policy suggestions that follow are drawn from the experiences reported in the Country Background Reports, the analyses of external review teams in Country Reviews and the available research literature. It should be stressed that there is no single model or global best practice of school leader appraisal. The development of practices always needs to take into account country-specific traditions and features of the respective education systems (also see Chapter 3). Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing policies for school leader appraisal for different countries. The implications also need to be treated with caution as, in some instances, further research is needed for specific contexts as well as across a sufficient number of countries to be confident about the consequences of specific approaches. Further potential areas of research on school leader appraisal include, among others, school leaders' perceptions of using separate appraisal processes for developmental and accountability purposes; the effects of using professional school leadership standards as a reference standard for appraisal; the effects of involving peer evaluators in the appraisal process; the effects of using teacher, parent and student surveys and questionnaires as a source of information; ways to strengthen links between appraisal and professional development; and the effects of using appraisal results to reward successful school leaders with a financial reward.

Governance

Promote the effective appraisal of school leaders within the broader assessment and evaluation framework while giving room to local diversity

Given the key role that school leaders play for teaching and learning, it should be a policy priority to strengthen processes for school leader appraisal. These need to reflect country-specific governance frameworks, the allocation of responsibilities in the education system and the extent of decentralisation. The existence of national curricula and standards and the overall culture of evaluation all need to be taken into account in approaches to the appraisal of school leaders (also see Chapter 3).

The development of a central/state policy framework for the systematic appraisal of school leaders constitutes one possible option to strengthen the appraisal of school leaders. Even though implementation is always key and measures need to be in place to ensure the successful implementation of appraisal (e.g. through building local capacity; through tools and support to school organising bodies to appraise school leaders; through ensuring ownership by school agents, etc.), a central/state framework can help ensure national consistency. It can contribute to ensuring that an appraisal process fulfils the required properties (e.g. validity, reliability, utility and fairness) and that appraisal is based on the latest research evidence. It can also help establish clear expectations among school leaders and evaluators about an appraisal process and help clearly map out the distribution of appraisal-related responsibilities among various actors (including education authorities at different levels of governance, intermediate authorities, school boards, school leaders, teachers, parents and students). Considering the importance of context for successful school leadership, central/state policy frameworks should leave sufficient scope to adjust procedures to local, school and individual circumstances (e.g. through the use of a range of reference standards, such as school development plans and individual job descriptions).

Policy makers also need to reflect on the ways in which to articulate school leader appraisal with the overall evaluation and assessment framework and its components to create synergies, to ensure that procedures are complementary and to avoid duplication and inconsistencies (e.g. in the communication of effective leadership, expected performance levels or a school leader's expected improvement, also see Chapter 3). In particular, policy makers need to consider the ways in which school leader appraisal is related to teacher appraisal and school evaluation. Many lessons of good practice from teacher appraisal may feed into the design and implementation of appraisal schemes for school leaders (also see Chapter 5). However, appraisal frameworks need to ensure that differences in teachers' and school leaders' responsibilities and required competencies are fully reflected in the appraisal regulations and that appraisal procedures are adjusted to provide effective support, feedback and learning opportunities for school leaders (e.g. through the practice of setting individual objectives). In countries with multiple procedures for the evaluation of the quality of school leadership, links between external school evaluation and individual school leader appraisal processes are key considering that both processes can usefully inform each other. Requirements or recommendations to use appraisal results in the external school evaluation process can help raise the importance of individual appraisal processes. Another option lies in requirements or recommendations for evaluators to draw on external school evaluation results as part of the individual school leader appraisal process. Expectations of school leadership that are communicated through external school evaluation as well as individual school leader

appraisal frameworks need to be coherent to communicate a clear and consistent picture of successful school leadership.

Clarify the purposes of school leader appraisal

Whichever the particular balance between accountability and developmental functions, appraisal ultimately needs to be contextualised within and contribute to a country's overall education objectives and the goal of improving teaching and learning for all students. This needs to be clearly communicated to everyone involved in the appraisal process. The frequent combination of accountability and developmental functions within a single appraisal process may lead to significant differences in the perceptions of the purposes of an appraisal between evaluators and school leaders and may risk undermining the effectiveness of an appraisal process. It is, therefore, essential to clearly communicate the particular purposes of an appraisal system and to ensure that these are shared by everyone involved. Policy frameworks need to clearly lay out what aspects of school leadership an appraisal seeks to evaluate, and for which purposes. While costs and benefits need to be weighed (e.g. in terms of workload for school leaders and evaluators), the introduction of separate processes for different purposes may constitute a further policy alternative.

Procedures

Develop a common leadership framework or set of professional standards for school leaders

A key element of school leader appraisal is a shared understanding of what counts as effective school leadership. A set of professional standards for school leaders that reflects the complexity of school leaders' tasks and responsibilities, that is informed by research evidence and involves school leaders in its development, can provide a clear and concise statement of the core elements of successful leadership by mapping out what school leaders are expected to know, be able to do, and how (see Ingvarson et al., 2006, for a literature review on the features and development of effective school leadership standards). School leadership frameworks or standards can guide the appraisal of school principals and deputy school principals or the appraisal of all types of school leadership positions. Developmental standards or frameworks, in particular, can distinguish between different levels of experience, development needs and leadership positions on a multilevel career structure.

Considering the need to balance a central frame of reference with local agency and contextualisation, leadership frameworks and professional standards should not be seen as a template or checklist against which school leaders are to be appraised, but rather as a reference point for the definition of individual objectives and/or the selection of appraisal aspects and criteria. The development of a central set of leadership standards that is open for adoption and adaptation by local agents may provide a further possibility to keep central leadership frameworks relevant to local needs.

In addition to appraisal processes, professional standards need to inform further key elements of a school leader's career, from selection and recruitment processes and initial school leadership preparation and induction programmes, to ongoing in-service training and professional development opportunities and career advancement.

Promote the appraisal of pedagogical leadership together with scope for local adaptation

Research has emphasised the need to select, prioritise and weight a select number of appraisal aspects and criteria that are precise, achievable, and measurable. In light of research on effective school leadership and some evidence on the effects of appraisal on leadership practices through the aspects that are assessed, school leader appraisal for improved student outcomes ideally focuses on the appraisal of practices and behaviours that research has identified as the core of pedagogical leadership (e.g. through professional school leadership standards). Appraisal, if well designed and implemented, can, then, reinforce the core objective of schools, i.e. high-quality teaching and learning. A focus on pedagogical leadership is also essential to encourage school leaders to take direct responsibility for the quality of learning and teaching in their school.

However, considering the role of local contexts, it is important that evaluators have an awareness that one set of school leadership practices and behaviours is not universally effective for all schools and at all times. Successful school leadership also depends on a school leader's choice on which areas to spend their time and efforts, and when. More management and administration-oriented tasks may, at times, be equally important as more pedagogical leadership tasks. The appraisal of a core set of leadership practices that form the basis of pedagogical leadership, therefore, needs to be balanced with scope for local flexibility and adaptation to local contexts. Scope for the local selection of appraisal aspects and criteria in line with central/state guidance that emphasise the importance of pedagogical leadership and/or the collaborative setting of objectives at a local level may help make appraisal manageable and relevant for local contexts. It may allow evaluators and school leaders to focus on priority areas relevant to a particular school and the leadership required in that context and analyse these in greater depth. However, considering the frequent combination of appraisal's functions for accountability and development, it is essential that policies ensure that the necessary selection of appraisal aspects and criteria reflects these purposes and individual as well as school needs (e.g. through the mandatory use of a range of reference standards and documents, such as individual job descriptions and school development plans).

Research indicates that school leaders in many contexts could focus more on their pedagogical leadership. Therefore, it is also important that school leaders benefit from ongoing support to develop pedagogical leadership skills through high-quality, targeted and relevant professional development opportunities before being held accountable for pedagogical leadership. Embedding appraisal for pedagogical leadership within a comprehensive leadership development framework would ensure school leaders are given the opportunity to develop the relevant knowledge and competencies to exercise pedagogical leadership. Appraisal focussed on pedagogical leadership can itself provide an opportunity for feedback and help identify areas for school leader's development. A comprehensive model for school leadership development would ensure strong linkages between appraisal and professional development.

Promote the appraisal of school leaders' competencies for monitoring, evaluation and assessment

School leaders play a key role for the effectiveness of evaluation and assessment frameworks, particularly for teacher appraisal and school evaluation (see Chapters 3, 5 and 6). In most countries, teacher appraisal involves the school principal or other school leaders, even though this involvement differs between and within countries. In countries

that have teacher appraisal procedures in place, this almost always entails some form of annual formal meeting between a school leader and a teacher. School leaders are also often one of the key figures in school evaluation processes. School leaders are often responsible for managing, stimulating and ensuring the effective implementation of school self-evaluation processes. They are also often responsible for ensuring good collaboration during external evaluations and for the adequate follow-up on the results of external evaluations including the development and implementation of school improvement plans. Practices related to monitoring, evaluation and appraisal (e.g. supporting and observing teachers, and observing students and classrooms), furthermore, constitute fundamental elements of pedagogical leadership focussing school leaders' attention towards teaching and learning.

Considering school leaders' key role in evaluation and assessment, appraisal should pay adequate attention to school leaders' responsibilities in this area. School leaders' essential role in internal teacher appraisal processes should be reflected in the appraisal of school leaders (e.g. through evaluating school leaders' competencies to manage staff; to authentically evaluate teaching and learning; to understand, observe and recognise good teaching; and to give developmental feedback to teachers). School leaders should equally be held directly accountable for and receive feedback on their ability to lead their school's self-evaluation processes, for their school's collaboration during external evaluations, and for the communication of external evaluation results to their school community.

However, it is equally important to provide school leaders with the opportunities to develop the necessary competencies to appraise teachers, lead school self-evaluation processes and get involved in external school evaluations (e.g. through school leadership development programmes that include aspects such as how to observe classrooms and interview teachers; how to analyse data; how to use school evaluation results; how to develop school improvement plans; how to involve teachers, students and parents in school self-evaluation, also see Chapters 5 and 6). Professional school leadership standards or frameworks also need to clarify and highlight the importance of school leadership for evaluation and assessment, in general, and for teacher appraisal and school evaluation, in particular.

Consider school leaders' efforts to distribute leadership, to enhance teacher leadership within schools, and to assume leadership responsibilities beyond their school borders as an integral part of appraisal

Research points to a range of advantages of leadership that is not just concentrated in one person, but distributed across several leadership staff in a school, including teachers, for sustained school improvement (Pont et al., 2008a; Schleicher, 2012). To give an example, distributed leadership in secondary schools can help provide regular informal feedback to teachers in their subject areas, something an individual school leader might not necessarily be equipped for considering the lack of expertise in different subjects (OECD, 2009). At the same time, distributing leadership in schools does not necessarily decrease a school leader's workload and may create new challenges for school leaders, who, in turn, require ongoing support.

Considering the role of appraisal for communicating a shared understanding of effective school leadership and the potential impact of appraisal on school leaders' practices and behaviours, appraisal should examine the ways in which school leaders foster distributed leadership in their schools as one aspect of appraisal (e.g. school leaders' competencies for building structural capacity, school leaders' efforts to create

opportunities for teacher leadership, school leaders' ability to enhance their teaching staff's capacity to lead, school leaders' ability to foster succession planning). Appraisal may constitute an important channel for providing feedback on the arrangements of distributed leadership. It may help inform professional development and wider support structures. It may also provide an opportunity to provide feedback to school leaders on their efforts to enhance teacher leadership in their schools. However, evaluators need to take barriers for the effective distribution of leadership (e.g. legal regulations and a lack of resources) into account.

Appraisal procedures should also reflect the growing importance of leadership tasks beyond school borders as a way of sharing expertise for system-wide improvement. Appraisal constitutes one opportunity for giving feedback on school leader's efforts to assume wider leadership tasks and responsibilities and can be used to point school leaders towards opportunities to engage in activities that may help improve the wider system.

Promote the use of multiple instruments and sources of evidence

Effective school leader appraisal requires the right mix and number of instruments and sources of information to judge a school leader's performance according to identified appraisal aspects and criteria. Research has increasingly stressed the benefits of using multiple tools to form a fair, valid and reliable picture of a school leader's performance from a comprehensive perspective.

The local selection of relevant sources of information by school leaders and evaluators may increase the relevance of the information that is gathered and create trust, acceptance and collaboration among school leaders. However, school leaders and evaluators need to have the necessary capacity and competence to choose suitable instruments and understand fairness, reliability and validity concerns.

Limited research has provided some insights into the benefits of different tools and the caution needed when using others:

- The use of school leader portfolios, if embedded within wider support structures, may ensure a school leader's views are adequately represented in the appraisal process and help strengthen the formative dimension of appraisal.
- The use of stakeholder surveys requires an awareness among evaluators of the politics that appraisal may involve. Teachers' views may add most value to an appraisal process considering their close insights into a school's daily routine.
- Given the wide range of factors that influence student outcomes within and outside schools, and persistent evidence that the impact of school leaders on student learning is mainly indirect and mediated through others, holding school leaders directly accountable for improved student test scores or the value-added by the teachers in their school faces serious challenges and risks.

Capacity

Build capacity for effective school leader appraisal

The successful implementation of any school leader appraisal system greatly depends on the selection and in-depth training of the evaluators. Policy makers, therefore, need to pay adequate attention to ensuring the credibility and competences of those evaluating

school leaders, especially considering the high levels of local autonomy for school leader appraisal in many contexts.

For building capacity for implementing central/state frameworks at a local level, funding for evaluators' training and professional development, the piloting of newly developed appraisal systems before implementation, opportunities for on-site formal training sessions for evaluators to discuss their experiences and the development of online platforms for ongoing discussion constitute possible measures to promote and develop appraisal-related skills among evaluators.

School leaders themselves also need to be provided with support to understand the appraisal purposes and procedures and to be competent in the use of results to benefit from the appraisal process. School leader preparation and professional development programmes should ensure that school leaders are aware of expectations and the meanings of effective school leadership that they are appraised against. School leaders need preparation to make the best use of the feedback received. The development of appraisal-related competencies should, therefore, constitute an important aspect of school leadership preparation programmes and ongoing professional development.

Promote school leader appraisal as an opportunity for peer learning

School leadership can be a lonely role and school leaders can face feelings of professional isolation. In this context, appraisal can provide an essential source of support for school leaders. Appraisal can help create opportunities for school leaders to learn from each other and to share good practices with school leaders from other schools. This may involve school leaders appraising other leaders as critical friends that face similar challenges and work conditions. While the costs in terms of the additional workload for school leaders need to be taken into account, the involvement of school leaders as one part of an evaluator team in the appraisal of other school leaders can strengthen the formative dimension of appraisal. The involvement of school leaders in each other's appraisal can, ultimately, also strengthen school leadership beyond the borders of a single school. Such collaboration and peer learning with other schools is an essential part of pedagogical leadership and may help foster system-wide improvement.

Furthermore, including a peer-appraisal element in the appraisal process may have beneficial effects on the appraisal process itself. Involving school leaders as peer-evaluators has the potential to increase the credibility and acceptance of the evaluator team among school leaders. Participation of school leaders in each other's appraisal can also help build capacity and develop a deeper understanding of the appraisal system among school leaders.

Use of results

Ensure school leader appraisal informs professional development

Given school leaders' crucial role for teaching and learning, it is important that they benefit from dedicated training and professional development opportunities. Appraisal is unlikely to produce effective results if it is not appropriately linked to professional development. Considering that the appraisal of school leaders involves the risk of increasing school leaders' workload and stress levels, it is essential to develop and successfully implement appraisal procedures that school leaders themselves perceive as meaningful and useful for improving their practices and behaviours.

Mechanisms should be in place to ensure that school leader appraisal itself provides effective and useful feedback that school leaders can use for improving their practices and behaviours (e.g. through a continuous cyclical appraisal process and the regular interaction between evaluators and appraisees). Appraisal procedures should, furthermore, feed into professional development activities and result in the preparation of an individual development plan. The formulation of steps for professional development should take into account the overall school objectives and the school development plan to ensure that appraisal and professional development are targeted towards the needs of a particular school. Considering the need to strengthen pedagogical leadership, individual development plans and professional development opportunities should pay adequate attention to developing school leaders' competencies and practices in this area. Appraisal can constitute one opportunity to identify further development needs to this end. Including professional development activities as one aspect and criterion of appraisal in the process provides a further possibility to strengthen ties between appraisal and professional development.

Consider the development of a career structure and career advancement opportunities to reward successful school leaders

It is essential to bear the contemporary challenges for school leadership in mind. The OECD project on Improving School Leadership by Pont et al. (2008a) highlighted the impact of a heavy workload coupled with a lack of adequate support and remuneration and uncertain career advancement prospects as some of the reasons for a lack of attracting talented new school leaders. As the OECD project on school leadership suggested, career development prospects as well as salary scales for school leaders that are separate from teachers' salary scales and that reflect leadership structures and school-level factors may help attract high performing leaders to all schools (Pont et al., 2008a).

In many countries, the absence of career opportunities for effective school leaders may undermine the role of appraisal. Using appraisal results to inform career advancement may help make appraisal for accountability more effective and meaningful. It is, however, important to bear in mind that research on the effects of such systems that tie appraisal to career advancement is scarce.

If appraisal is linked to school leaders' career advancement, policy makers need to ensure that sound, valid and reliable appraisal processes are in place that school leaders perceive as fair and objective. This requires reliable indicators and clear appraisal aspects and criteria, training for evaluators and due consideration for the context in which a school leader works.

Notes

1. Recent major research studies on the relationship between school leadership and student outcomes and the nature of effective school leadership include Day et al. (2009), Robinson et al. (2009), and Louis et al. (2010). Further research undertakings include work of the European project The Making of: Leadership in Education to develop a European Framework of Reference on Effective Leadership (*www.leadership-in-education.eu*), the European Leadership Improvement for Student Achievement project (LISA) (*www.leadership-in-education.eu/index.php?id=235*) and the International Successful School Principalship Project (*www.uv.uio.no/ils/english/research/projects/isspp*).

2. In the research literature, pedagogical leadership is often also referred to as learning-centred leadership and leadership for learning.

3. In the research literature, distributed leadership is, at times, also referred to as collaborative leadership and shared leadership, among others.

4. Chile provides an example of an appraisal scheme that is specifically designed to encourage the successful distribution of leadership in schools through leadership teams (see Box 7.2).

5. Following previous work on school leadership undertaken by the OECD Directorate for Education and Skills, and for the sake of greater coherence across the chapter, this chapter refers to school leaders, school principals and deputy school principals. Depending on the country context, this refers to concepts such as the school principal in Australia, Canada and the United States or the headteacher in the United Kingdom and Ireland. However, depending on the context of a particular research study cited, different terms may be used. For instance, reference to a study on the effects of school leadership in the United Kingdom may describe the highest leadership position in schools as the headteacher.

6. In Mexico, school principals and pedagogical-technical advisors in schools at ISCED levels 1 and 2 are appraised through the Universal Teacher Appraisal System and the National Teaching Career Programme (*Programa Nacional de Carrera magisterial* [PNCM]) (see Chapter 5). The appraisal of school leaders through the PNCM system is based on information on student academic achievement (e.g. in standardised central assessments ENLACE) (50%), a school leader's professional development activities (20%), and his or her exercise of tasks and responsibilities (30%). The assessment of a school leader's exercise of tasks and responsibilities includes a test on professional preparation (5%). The appraisal framework described for Mexico throughout this chapter refers to the exclusive appraisal scheme for school principals at centrally managed public schools at ISCED level 3.

7. The description of appraisal procedures, aspects and criteria, and instruments and sources of information focuses on the internal element of the school leader appraisal system in Portugal. For further details on the external element (external school evaluation), see Chapter 6.

8. Leithwood et al.'s (2004) conclusions are based on a review of quantitative and qualitative research. This review formed the basis for a six year quantitative and qualitative research study by Louis et al. (2010). Louis et al. collected data from a wide range of respondents in the United States across nine states, 43 school districts, and 180 primary, middle and secondary schools. This amounted to survey data from a total of 8 391 teachers and 471 school administrators; interview data from 581 teachers and administrators, 304 district level informants, and 124 state personnel; and observational data from 312 classrooms. Louis et al. also obtained student achievement data for literacy and mathematics in primary and secondary years, using scores on the states' tests for measuring Adequate Yearly Progress as mandated by the No Child Left Behind Act of 2002.

9. Day et al.'s (2009; 2010) research was based on a sample of schools in England that had improved student outcomes over at least three consecutive years under the leadership of the same headteacher. It included surveys completed by headteachers and a range of other stakeholders, 20 case studies of primary and secondary schools and a literature review.

10. Emstad's case study focussed on two primary schools that implemented a formative self-evaluation process that involved staff, students and parents. The self-evaluation process aimed to increase student learning through the use of student assessment and clear learning goals and relied on the same tool in both schools. The evidence was based on interviews with the school principal, three teachers (Years 5, 6 and 7) and a group of students at each school and additional documents and observations.

11. Sun and Youngs examined school principal appraisal and practices in 13 school districts in Michigan, United States with a focus on the relationships between school principals' behaviours and the appraisal purposes, focus and assessed leadership practices and behaviours. The study relied on data on district enrolment size in 2006-07, data on the percentage of students in each district eligible for free or reduced-price lunch in 2006-07 and a survey of district administrators, school principals and teachers (including 19 administrators responsible for the appraisal of primary and/or secondary school principals in 2006-07).

12. Sun et al.'s study examined school principal appraisal and its relation to school leadership practices and behaviours in Michigan, United States, and Beijing, China. Based on survey data of two independent samples in Michigan (2007-08) and Beijing (2008-09), Sun et al. developed a two-level Multivariate Hierarchical Linear Model. The survey included 88 primary and/or secondary school principals in Michigan and 90 primary and/or secondary school principals in metropolitan Beijing and similar demographic characteristics.

13. Parylo et al.'s study was based on semi-structured interviews with 16 school principals from 4 school districts in the state of Georgia, United States. The sample included school principals form all school levels and an even ratio between male and female school principals at different stages of their career. The school districts were selected based on their size, Adequate Yearly Progress status, socio-economic status and location. All of the school districts used different appraisal tools, but all of the procedures involved a pre-observation, an observation and a final appraisal meeting.

14. For further information on Chile's School Principals of Excellence training programme, see *www.formaciondirectores.mineduc.cl.* Also see *www.gestionescolar.cl* for further reference.

15. For further information on New Zealand's school leadership initiatives, see *www.educationalleaders.govt.nz*.

16. For further information on Sweden's National School Leadership Training Programme, see *www.skolverket.se/publikationer?id=2254.*

17. Italy and the Netherlands have been developing central frameworks for school leader appraisal.

18. After two decades of decentralisation, Hungary has experienced a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities (including funding allocation) from 2013 onwards. The information about the distribution of responsibilities for the appraisal of school leaders in this chapter refers to the period prior to this reform. The Hungarian appraisal system is envisaged to change substantially in 2013.

19. Ultra-orthodox religious schools are exempt from central requirements for school leader appraisal.

20. Thomas et al.'s (1999) study involved 66 school districts and superintendents and 62 school principals in the province of Alberta, Canada.

21. Gaziel's (2008) study involved semi-structured interviews with two primary school supervisors and six primary school principals chosen randomly from four of the six educational districts in Israel. Israel introduced a new framework for the appraisal of school leaders in primary schools and partially in lower secondary schools in 2010-2011. Gaziel's research refers to appraisal procedures in place prior to the implementation of the new appraisal model.

22. Davis and Hensley's (1999) study involved a review of relevant literature as well as in-depth interviews with six superintendents and 14 school principals in primary and secondary schools located in urban, suburban and rural school districts in California, United States.

23. For further information, see *www.education.vic.gov.au/school/principals/profdev/Pages/principalpd.aspx*

24. For further information, see *http://edu.gov.on.ca/eng/policyfunding/leadership/appraise.html*

25. For further information, see *El Portal Gestión y Liderazgo Educativo* (*www2.gestionyliderazgoeducativo.cl/gestioncalidad/evaluacion/home/index.php*) and *Más Directivos* (*www.masdirectivos.cl*).

26. For further information, see *www.minedu.govt.nz/Boards/ManagingResources/PerformanceManagement.aspx* and *www.nzsta.org.nz/board-as-employers/principal/principals-appraisal/*

27. The Portuguese school leader appraisal system relies on a quota system for the top two rating levels at the national level, which is monitored by a national Appraisal Co-ordination Council.

28. At the time of printing this report, the Chilean Ministry of Education was planning to update the Framework for Good School Leadership and was developing more specific professional standards for school principals.

29. The portfolio approach was, ultimately, not adopted for wider usage by the state of Ohio.

30. The appraisal scheme that Israel's central education authorities introduced for school principals from 2010-11 onwards does not include student achievement data as part of the appraisal process. The information reported by Israeli school principals as presented refers to the time before implementation of the new system and relates mainly to informal appraisal processes at a local level.

31. The Regional Training Unit is a regional school leadership development organisation that caters for the needs of school leaders at every stage of their careers. It was established in 1990 to: 1) provide elements of the Education and Library Boards' training programmes which can most effectively be undertaken on a regional basis; 2) undertake responsibility for the long-term management training of principals and senior staff of schools; and 3) provide training for school governors and for Education and Library Boards' staff and its members. The arrangements for professional support are envisaged to transfer to the new Education and Skills Authority from 2013 onwards (Department of Education, Northern Ireland (forthcoming).

32. For further information, see the following websites: *http://lds21.com* and *www.leadership-in-education.eu/index.php?id=40*.

33. This information reflects the situation prior to the introduction of new legislation in 2011 and 2012 that has led to greater degrees of central decision-making in education.

34. For further information on the Pittsburgh Urban Leadership system for Excellence (PULSE), see *www.pps.k12.pa.us/1431104301323045O/site/default.asp*.

References

Amsterdam, C.E., R.L. Johnson, D.E. Monrad and S.L. Tonnsen (2003), "A collaborative approach to the development and validation of a principal evaluation system: A case study", *Journal of Personnel Evaluation in Education,* Vol. 17, No. 3, pp. 221–242.

Asia Society (2012), *Teaching and Leadership for the Twenty-First Century – The 2012 International Summit on the Teaching Profession*, report prepared by the Asia Society, *http://asiasociety.org/files/2012teachingsummit.pdf*.

Bal, J. and J. de Jong (2007), *Improving School Leadership: OECD Review Background Report for the Netherlands*, report prepared for the Ministry of Education, Culture and Science in the context of the OECD activity "Improving School Leadership", Zoetermeer, Netherlands.

Bennett, N., C. Wise, P.A. Woods and J.A. Harvey (2003), *Distributed Leadership: A Review of Literature*, National College for School Leadership, Nottingham.

Brejc, M., M. Sardoč and D. Zupanc (2011), *Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Slovenia,* National School for Leadership in Education in cooperation with the National Examinations Centre, *www.oecd.org/edu/evaluationpolicy*.

Brown-Sims, M. (2010), *Evaluating School Principals*, National Comprehensive Center for Teacher Quality, Washington, DC, *www.tqsource.org/publications/KeyIssue_PrincipalAssessments.pdf*.

Bush, T. and D. Glover (2003), *School Leadership: Concepts and Evidence*, National College for School Leadership, Nottingham.

Catano, N. and J.H. Stronge (2007), "What do we expect of school principals? Congruence between principal evaluation and performance standards", *International Journal of Leadership in Education*, Vol. 10, No. 4, pp. 379-399.

CEPPE (Centro de Estudios de Políticas y Prácticas en Educación) (forthcoming), "Learning standards, teaching standards and standards for school principals: A comparative study", paper commissioned by the OECD and developed within the framework of the OECD-Mexico Co-operation Agreement to Improve the Quality of Education in Mexican Schools, CEPPE, Santiago de Chile.

Chilean Ministry of Education (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Chile*, forthcoming at *www.oecd.org/edu/evaluationpolicy.*

Claudet, J. (2002), "Integrating school leadership knowledge and practice using multimedia technology: Linking national standards, assessment, and professional development", *Journal of Personnel Evaluation in Education*, Vol. 16 (1), pp. 29-43.

Clifford, M. and S. Ross (2011), *Designing Principal Evaluation Systems: Research to Guide Decision-Making: An Executive Summary of Current Research*, paper prepared

in collaboration with the National Association of Elementary School Principals, American Institutes for Research, Washington DC, *www.naesp.org/sites/default/files/PrincipalEvaluation_ExecutiveSummary.pdf*.

Coffield, F. (2012), "Why the McKinsey reports will not improve school systems", *Journal of Education Policy*, Vol. 27, No. 1, pp. 131-149.

Condon, C. and M. Clifford, M. (2010), *Measuring Principal Performance: How Rigorous are Commonly Used Principal Performance Assessment Instruments?*, A Quality School Leadership Issue Brief, Learning Point Associates, Naperville, Illinois, *www.learningpt.org/pdfs/QSLBrief2.pdf*.

Crawford, M. and P. Earley (2004), "Headteacher performance management: An investigation of the role of the external adviser", *School Leadership and Management*, Vol. 24, No. 4, pp. 377-389.

Cullen, K. (1997), "An evaluation of the United Kingdom's national system of headteacher appraisal", *Studies in Educational Evaluation*, Vol. 23, No. 2, pp. 103-130.

Danish Ministry of Education and Rambøll (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Update to the Country Background Report for Denmark*, National Agency for Quality and Supervision, *www.oecd.org/edu/evaluationpolicy*.

Darling-Hammond, L. and R. Rothman (2011), "Lessons learned from Finland, Ontario, and Singapore", in L. Darling-Hammond and R. Rothman (eds.), *Teacher and Leader Effectiveness in High-Performing Education Systems*, Alliance for Excellent Education, Washington DC, Stanford Center for Opportunity Policy in Education, Stanford, California, *http://edpolicy.stanford.edu/sites/default/files/publications/teacher-and-leader-effectiveness-high-performing-education-systems.pdf*.

Davis, S.H. and P.A. Hensley (1999), "The politics of principal evaluation", *Journal of Personnel Evaluation in Education*, Vol. 13, No. 4, pp. 383-403.

Davis, S., K. Kearney, N. Sanders, C. Thomas and R. Leon (2011), *The Policies and Practices of Principal Evaluation: A Review of the Literature*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1104.pdf*.

Day, C., P. Sammons, D. Hopkins, A. Harris, K. Leithwood, Q. Gu and E. Brown (2010), *10 Strong Claims about Successful School Leadership*, National College for School Leadership, Nottingham.

Day, C., P. Sammons, D. Hopkins, A. Harris, K. Leithwood, Q. Gu, E. Brown, E. Ahtaridou and A. Kington (2009), *The Impact of School Leadership on Pupil Outcomes: Final Report*, University of Nottingham, Nottingham.

Department of Education, Northern Ireland (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Northern Ireland, United Kingdom*, *www.oecd.org/edu/evaluationpolicy*.

Dos Santos, S. and R. Rakocevic (2012), *Les cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude de l'OCDE : Rapport national de base pour la France*, Direction de l'Évaluation, de la Prospective et de la Performance (DEPP), Ministère de l'Éducation nationale, *www.oecd.org/edu/evaluationpolicy*.

Duke, D.L. (1992), "Concepts of administrative effectiveness and the evaluation of school administrators", *Journal of Personnel Evaluation in Education*, Vol. 6, No. 2, pp. 103-121.

EACEA (Education, Audiovisual and Cultural Executive Agency) (2012), *Key Data on Education in Europe 2012*, EACEA, Brussels.

Emstad, A.B. (2011), "The principal's role in the post-evaluation process: How does the principal engage in the work carried out after the schools self-evaluation?", *Educational Assessment, Evaluation and Accountability*, Vol. 23, No. 4, pp. 271-288.

Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, *www.oecd.org/edu/evaluationpolicy*.

Fournier, G. and D. Mildon (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Canada*, Council of Ministers of Education, Canada (CMEC).

Gaziel, H. (2008), "Principals' performance assessment: Empirical evidence from an Israeli case study", *Educational Management Administration & Leadership*, Vol. 36, No. 3, pp. 337-351.

Ginsberg, R. and B. Berry (1990), "The folklore of principal evaluation", *Journal of Personnel Evaluation in Education*, Vol. 3, pp. 205-230.

Ginsberg, R. and T. Thompson (1992), "Dilemmas and solutions regarding principal evaluation", *Peabody Journal of Education*, Vol. 68, No.1, pp. 58-74.

Glasman, N.S. and L.D. Glasman (2010), "Principal evaluation: Concepts, needs and realities", *International Encyclopaedia of Education*, 3rd Edition, pp. 698-705.

Goldring, E., C.C. Cravens, J. Murphy, S.N. Elliott, B. Carson and A.C. Porter (2008), "The evaluation of principals: What and how do states and districts assess leadership?", paper presented at the Annual Meeting of American Educational Research Association, New York, March 2008.

Goldring, E., A.C. Porter, J. Murphy, S.N. Elliott and X. Cravens (2007), *Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices*, a report prepared for the Wallace Foundation grant on Leadership Assessment, March 2007, *www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/Assessing-Learning-Centered-Leadership.pdf*.

Grossman, T. (2011), *Preparing Principals to Evaluate Teachers*, Issue Brief, NGA Center for Best Practices, Washington DC, *www.wallacefoundation.org/knowledge-center/school-leadership/state-policy/Documents/Issue-Brief-Preparing-Principals-to-Evaluate-Teachers.pdf*.

Hajdúková, V., M. Šnídlová, A. Vykydalová, T. Lukács, A. Mentel, M. Pokorný, Z. Juščáková and R. Kanovská (forthcoming), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Slovak Republic,* National Institute for Certified Educational Measurements (NÚCEM), Bratislava, *www.oecd.org/edu/evaluationpolicy*.

Halász, G. (2009), "School leadership and leadership development in five Central European countries – synthesis report", in M. Schratz, L. Sláviková, E. Křížková, E. Kecliková, M. Szabó, E. Révész, V. Laššák, A. Hašková, J. Erčulj, P. Peček and G. Halász (eds.), *The Role of School Leadership in the Improvement of Learning. Country Reports and Case Studies of a Central-European Project*, Tempus Public Foundation, Budapest, pp. 185-267.

Hallinger, P. and R.H. Heck (1998), "Exploring the principal's contribution to school effectiveness: 1980-1995", *School Effectiveness and School Improvement*, Vol. 9, No. 2, pp. 157-191.

Harris, A. and J.P. Spillane (2008), "Distributed leadership through the looking glass", *Management in Education*, Vol. 22, No. 31, pp. 31-34.

Heck, R.H. and G.A. Marcoulides (1992), "Principal assessment: Conceptual problem, methodological problem, or both?", *Peadbody Journal of Education*, Vol. 68, No. 1, pp. 124-144.

Hellawell, D. and N. Hancock (1998), "Some benefits of primary headteacher appraisal: An experience potentially too good to lose?", *School Leadership and Management*, Vol. 18, No. 2, pp. 213-229.

Hopkins, D. (2008), "Realising the potential of system leadership", in B. Pont, D. Nusche and D. Hopkins (eds.), *Improving School Leadership, Volume 2: Case Studies on System Leadership*, OECD Publishing, Paris.

IIE (Institute for Information on Education) (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Czech Republic*, *www.oecd.org/edu/evaluationpolicy*.

Ingvarson, L., M. Anderson, P. Gronn and A. Jackson (2006), *Standards for School Leadership: A Critical Review of Literature*, Teaching Australia, Australian Institute for Teaching and School Leadership, Melbourne, *http://research.acer.edu.au/cgi/viewcontent.cgi?filename=0&article=1002&context=teaching_standards&type=additional*.

Johnston, M. and M. Thomas (2005), "Riding the wave of administrator accountability: A portfolio approach", *Journal of Educational Administration*, Vol. 43, No. 4, pp. 368-386.

Keating, B. (2011), *How Four Districts Crafted Innovative Principal Evaluation Systems: Success Stories in Collaboration*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1109.pdf*.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and Jungha Park (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

Kimball, S.M., A. Milanowski and S.A. McKinney (2009), "Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial", *Leadership and Policy in Schools*, Vol. 8, No. 3, pp. 233-263.

Krüger, M. and J. Scheerens (2012), "Conceptual perspectives on school leadership", in J. Scheerens (ed.), *School Leadership Effects Revisited: Review and Meta-Analysis of Empirical Studies*, SpringerBriefs in Education, Springer, Dordrecht, Heidelberg, London, New York, pp. 1-31.

Lashway, L. (2003), "Improving principal evaluation", *ERIC Digest 172*, ERIC Clearinghouse on Educational Management, Eugene, Oregon.

Leithwood, K. (2001), "School leadership in the context of accountability policies", *International Journal of Leadership in Education*, Vol. 4, No. 3, pp. 217-235.

Leithwood, K., C. Day, P. Sammons, A. Harris and D. Hopkins (2006), *Successful School Leadership: What it is and How it Influences Pupil Learning*, National College for School Leadership, Department for Education and Skills, Nottingham, *www.education.gov.uk/publications/eOrderingDownload/RR800.pdf*.

Leithwood, K., K.S. Louis, S. Anderson and K. Wahlstrom (2004), *How Leadership Influences Student Learning*, University of Minnesota, University of Toronto, a report commissioned by the Wallace Foundation, *www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf*.

Leon, R., S. Davis, K. Kearney, N. Sanders and C. Thomas (2011), *Key Features of a Comprehensive Principal Evaluation System*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1107.pdf*.

Louis, K.S., K. Leithwood, K.L. Wahlstrom and S.E. Anderson (2010), *Investigating the Links to Improved Student Learning – Final Report of Research Findings*, University of Minnesota, University of Toronto, a report commissioned by the Wallace Foundation, *www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf*.

Louis, K.S. and V.M. Robinson (2012), "External mandates and instructional leadership: School leaders as mediating agents", *Journal of Educational Administration*, Vol. 50, No. 5, pp. 629-665.

Marzano, R.J., T. Waters and B.A. McNulty (2005), *School Leadership that Works: From Research to Results*, McRel, Aurora, Colorado.

Matthews, P., H. Moorman and D. Nusche (2008), "Building leadership capacity for system improvement in Victoria, Australia", in B. Pont, D. Nusche and D. Hopkins (eds.), *Improving School Leadership Volume 2: Case Studies on System Leadership*, OECD Publishing, Paris.

Mattson Almanzán, H., N. Sanders and K. Kearney (2011), *How Six States Are Implementing Principal Evaluation Systems*, WestEd, San Francisco, California, *www.wested.org/online_pubs/resource1105.pdf*.

Mulford, B. (2008), "The leadership challenge: Improving learning in schools", *Australian Education Review*, No. 53, Australian Council *for* Educational Research, Camberwell, Victoria, *http://research.acer.edu.au/aer/2/*.

Mulford, B. (2003), "School leaders: Changing roles and impact on teacher and school effectiveness", paper commissioned by the OECD Education and Training Policy Division for the activity Attracting, Developing and Retaining Effective Teachers, *www.oecd.org/dataoecd/61/61/2635399.pdf*.

Murphy, J., S.N. Elliott, E. Goldring and A.C. Porter (2006), *Learning-Centered Leadership: A Conceptual Foundation*, a report prepared for the Wallace Foundation Grant on Leadership Assessment, August 2006, *www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/Learning-Centered-Leadership-A-Conceptual-Foundation.pdf*.

New Leaders for New Schools (2010), *Evaluating Principals: Balancing Accountability with Professional Growth*, New Leaders for New Schools, New York, NY.

Normore, A.H. (2004), "The edge of chaos: School administrators and accountability", *Journal of Educational Administration*, Vol. 42 (1), pp. 55-77.

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, *www.oecd.org/edu/evaluationpolicy.*

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

OECD (2012), *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing, Paris.

OECD (2010), *PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV),* OECD Publishing, Paris.

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.

Ovando, M.N. and A. Ramirez Jr. (2007), "Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success", *Journal of Personnel Evaluation in Education*, Vol. 20, No. 1-2, pp. 85-110.

Parylo, O. (2012a), "Evaluation of educational administration: A decade review of research (2001-2010)", *Studies in Educational Evaluation*, Vol. 38, No. 3-4, pp. 73-83.

Parylo, O., S.J. Zepeda and E. Bengtson (2012b), "Principals' experiences of being evaluated: A phenomenological study", *Educational Assessment, Evaluation and Accountability*, Vol. 24, No. 3, pp. 215-238.

Pashiardis, P. and S. Brauckmann (2008), "Evaluation of school principals", in J. Lumby, G. Crow and P. Pashiardis (eds.), *International Handbook on the Preparation and Development of School Leaders*, Routledge, New York, London, pp. 263-280.

Pervin, B. and C. Campbell (2011), "Systems for teacher and leader effectiveness and quality: Ontario, Canada", in L. Darling-Hammond and R. Rothman (eds.), *Teacher and Leader Effectiveness in High-Performing Education Systems*, Alliance for Excellent Education, Washington DC, Stanford Center for Opportunity Policy in Education, Stanford, California.

Piggot-Irvine, E. (2003), "Key features of appraisal effectiveness", *The International Journal of Educational Management*. Vol. 17 (4), pp. 170-178.

Pont, B., D. Nusche and H. Moorman (2008a), *Improving School Leadership Volume 1: Policy and Practice*, OECD Publishing, Paris.

Pont, B., D. Nusche and D. Hopkins (2008b), *Improving School Leadership Volume 2: Case Studies on System Leadership*, OECD Publishing, Paris.

Porter, A.C., E. Goldring, J. Murphy, S.N. Elliott and X. Cravens (2006), *A Framework for the Assessment of Learning-Centered Leadership*", a report prepared for the Wallace Foundation grant on Leadership Assessment, September 2006, *www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/A-Framework-for-the-Assessment-of-Learning-Centered-Leadership.pdf*.

Portin, B.S., S. Feldman and M.S. Knapp (2006), *Purposes, Uses, and Practices of Leadership Assessment in Education*, Center for the Study of Teaching and Policy, University of Washington, Seattle, WA, *http://depts.washington.edu/ctpmail/PDFs/LAssess-Oct25.pdf*.

Portuguese Ministry of Education and Science (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Portugal*, *www.oecd.org/edu/evaluationpolicy*.

PriceWaterhouseCoopers (2009), *School Leadership Matters: An empirical assessment of the attractiveness of principalship in the North of Ireland and the South of Ireland*, final report prepared for the Regional Training Unit (RTU) and Leadership Development for Schools (LDS), *www.rtuni.org/uploads/docs/Teacher%20Ed%20-%20School%20Leadership%20Matters%20Report.pdf.*

PriceWaterhouseCoopers (2007), *Independent Study into School Leadership: Main Report*, Department for Education and Skills, London, England.

Radinger, T. (forthcoming), "School leader appraisal: Current practices and a literature review", *OECD Education Working Papers*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/workingpapers*.

Reeves, D.B. (2009), *Assessing Educational Leaders: Evaluating Performance for Improved Individual and Organizational Results*, 2nd Edition, Corwin Press, Thousand Oaks, California.

Robinson, V., M. Hohepa and C. Lloyd (2009), *School Leadership and Student Outcomes: Identifying What Works and Why*, New Zealand Ministry of Education, Wellington.

Sanders, N. and K. Kearney (2011), *A Brief Overview of Principal Evaluation Literature: Implications for Selecting Evaluation Models*, California Comprehensive Assistance Center, WestEd, San Francisco, California.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012a), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012b), *OECD Reviews of Evaluation and Assessment in Education: Mexico*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Schleicher, A. (ed.) (2012), *Preparing Teachers and Developing School Leaders for the 21st Century: Lessons from around the World*, OECD Publishing, Paris.

Scottish Executive (2005), *Ambitious, Excellent Schools: Leadership – A Discussion Paper*, Scottish Executive, Edinburgh.

Scottish Executive (2004), *Ambitious, Excellent Schools: Our Agenda for Action*, Scottish Executive, Edinburgh.

Spillane, J.P., R. Halverson and J.B. Diamond (2004), "Towards a theory of leadership practice: A distributed perspective", *Journal of Curriculum Studies*, Vol. 36, No. 1, pp. 3-34.

Spillane, J.P. and A.W. Kenney (2012), "School administration in a changing education sector: The US experience", *Journal of Educational Administration*, Vol. 50, No. 5, pp. 541-561.

State of Victoria (Department of Education and Early Childhood Development) (2012), *New directions for school leadership and the teaching profession Discussion paper June 2012*, Communications Division for Flagship Strategies Division Department of Education and Early Childhood Development, Melbourne, *www.eduweb.vic.gov.au/edulibrary/public/commrel/about/teachingprofession.pdf*.

Stufflebeam, D. and D. Nevo (1993), "Principal evaluation: New directions for improvement", *Peabody Journal of Education*, Vol. 68, No. 2, pp. 24-46.

Sun, M. and P. Youngs (2009), "How does district principal evaluation affect learning-centered principal leadership? Evidence from Michigan school districts", *Leadership and Policy in Schools*, Vol. 8, No. 4, pp. 411-445.

Sun, M., P. Youngs, H. Yang, H. Chu and Q. Zhao (2012), "Association of district principal evaluation with learning-centered leadership practice: Evidence from Michigan and Beijing", *Educational Assessment, Evaluation and Accountability*, published on line 11 April 2012.

Thomas, D.W., E.A. Holdaway and K.L. Ward (2000), "Policies and practices involved in the evaluation of school principals", *Journal of Personnel Evaluation in Education*, Vol. 14, No. 3, pp. 215-240.

Wallace Foundation (2012), *Wallace Perspective: The School Principal as Leader: Guiding Schools to Better Teaching and Learning*, Wallace Foundation, New York.

Wallace Foundation (2009), *Wallace Perspective: Assessing the Effectiveness of School Leaders: New Directions and New Processes,* Wallace Foundation, New York, *www.wallacefoundation.org/knowledge-center/school-leadership/principal-evaluation/Documents/Assessing-the-Effectiveness-of-School-Leaders.pdf*.

Annex 7.A Features of school leader appraisal frameworks

The tables below provide information on features of school leader appraisal frameworks in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. The information was supplied by countries through a questionnaire specifically developed by the OECD Review.

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Table 7.A.1 School leadership appraisal practices

This table describes central/state policy frameworks for the appraisal of school leaders, i.e. the external evaluation of individual school leaders to make a judgement about their work and performance using objective criteria. Appraisal results may be used to inform professional development, career advancement and rewards for school leaders.

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
Australia	School principal; deputy school principal[1]	Mandatory periodic (frequency varies nationally, but generally annually)	School principals and deputy school principals: state education authorities Deputy school principals only: school principals	School principals: state education authorities Deputy school principals: school principals	State standards for school leadership; a description of the general and professional duties of school leaders	General leadership; pedagogical leadership; organisation development; school climate; community relations; evaluation and accountability; resource management; interpersonal skills; school performance in relation to strategic plan	School principals: dialogue with evaluators; school leader self-appraisal; school visit; student outcomes; parent surveys Deputy school principals: dialogue with school principal; school leader self-appraisal; staff survey	Yes (varies nationally)	Yes, it systematically results in a professional development plan	Yes, appraisal results may influence access to career development opportunities	Permanent salary increment; extra professional development opportunities	Further appraisal; compulsory training; deferral of promotion; transfer; dismissal
Austria	None[2]	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.) (*)	School principal (public and government-dependent private schools only)[3]	Mandatory periodic (at least every 4 years)	Central education authority	School organising bodies[4]	Individualised job descriptions, adapted to local expectations[5]	At the discretion of school organising bodies	Dialogue with evaluators	Yes (2 levels)	Yes, it is expected to influence professional development activities	Yes, appraisal results may influence access to career development opportunities (but this is not defined in legislation)	None	Dismissal[6]
Belgium (Fr.)	School principal; deputy school principal (public and government-dependent private schools only)	Mandatory periodic (every 5 years)	Central education authority	Public schools: central education authority Government-dependent private schools: school board (may involve external experts)	A description of the general and professional duties of school leaders (*lettre de mission*); central standards for school leadership	Pedagogical leadership; interpersonal skills; resource management; additional aspects that are not specified[7]	At the discretion of evaluators (typically dialogue with evaluators)	No[8]	Varies across schools (appraisal may result in suggestions for professional development)	No	None	Further appraisal

Table 7.A.1 School leadership appraisal practices (*continued*)

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
Canada	Varies, at the discretion of provincial/territorial or local education authorities[9]	Varies, at the discretion of provincial/territorial authorities, local education authorities, school boards/committees	Provincial/territorial or local education authorities	Provincial/territorial or local education authorities	A description of the general and professional duties of school leaders; provincial/territorial standards for school leadership (e.g. Alberta and Ontario)	Pedagogical leadership; community relations; organisation development; resource management	Dialogue with evaluators; school visits	Varies across provinces (e.g. in Ontario school leaders' performance can be rated as satisfactory or unsatisfactory)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence the speed at which a school leader progresses in the career structure	None	Deferral of promotion; dismissal from leadership role
Chile	Performance appraisal (*Evaluación de Desempeño*, Law 19.979): teaching and technical-pedagogical leaders (including school principals, deputy school principals, general inspectors, and heads of technical units) (public schools only)[10]	Mandatory periodic (annually)	Central education authority or government	School principals: local education authorities (municipal sustainers) Other teaching and technical-pedagogical leaders: school principal	School Management Quality Framework (*Modelo de Calidad de la Gestión Escolar*) for institutional targets; national standards for school leadership (*Marco para la Buena Dirección*) for professional development objectives; school development plan; annual plan for the development of education of the local education authority (municipality)	Achievement of 2-4 institutional targets and 2-4 professional development objectives, related to general leadership; pedagogical leadership; school climate; and resource management[11]	Indicators and information sources selected by school leaders and evaluators compiled into an Evidence Portfolio (*Carpeta de Evidencia*); Implementation report (*Reporte de Implementación*)	Yes (rating on a scale 0-100 and 4 proficiency levels)	Yes, it is systematically linked with professional development activities	No	Salary increment for a fixed period of time (1-year period)	Compulsory training; dismissal[12]
	Appraisal of Collective Performance (*Asignación de Desempeño Colectivo*, laws 19.933 and 20.158): teaching and technical-pedagogical leadership teams (public schools and non-government dependent private schools with more than 250 students only)	Voluntary	Central education authority or government	Performance agreements submitted by school principals: school organising bodies (sustainers)[13] Performance agreements submitted by other teaching and technical-pedagogical leaders: school principals[13]	School Management Quality Framework (*Modelo de Calidad de la Gestión Escolar*) for institutional targets; school development plan; annual plan for the development of education of the local education authority (municipality) or school board	Achievement of a performance agreement that entails an institutional objective and institutional targets (*Convenio*), related to general leadership; pedagogical leadership; school climate; and resource managament[14]	Indicators and information sources selected by the school leadership team and evaluators compiled into an Evidence Portfolio (*Carpeta de Evidencia*); implementation report (*Reporte de Implementación*)	Yes (rating on a scale 0-100 and 4 proficiency levels)	No	No	One-off financial bonus	None

Table 7.A.1 School leadership appraisal practices (*continued*)

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
Czech Republic	School principals (public schools only)[15]	In relation to decision on employment status (contract renewal)	School organising bodies	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Yes	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body; underperformance may result in dismissal
Denmark	None[16]	a	a	a	a	a	a	a	a	a	a	a
Estonia	None[17]	a	a	a	a	a	a	a	a	a	a	a
Finland	None[18]	a	a	a	a	a	a	a	a	a	a	a
France	School principal; deputy school principal (public schools)[19]	Mandatory periodic (every 3 years); at the request of the school principal in the event of a transfer request	Central education authority (Ministry of Education)	ISCED level 1: local inspection bodies (IEN);[20] regional education authorities (*recteurs d'académies*) ISCED levels 2 and 3: regional education authorities (*recteurs d'académies*)	School principal's school analysis documents (*le diagnostic de l'établissement*); a description of the general and professional duties/mission statement (*lettre de mission*)	ISCED level 1: observation of leadership practices and behaviours (general leadership; pedagogical leadership; community relations; resource management) ISCED levels 2 and 3: achievement of objectives set by the *lettre de mission*; observation of leadership practices and behaviours (general leadership; pedagogical leadership; community relations; resource management)	ISCED level 1: dialogue with evaluators ISCED levels 2 and 3: dialogue with evaluators; indicators and information sources selected by school principals and evaluators	Yes (4 levels: exceptional; very good; good; to be improved)	No	Yes, appraisal results influence the speed at which a school leader progresses in the career structure or salary scale	Salary increment; one-off financial bonus	Deferral of promotion; salary increment withheld
Hungary	None[21]	a	a	a	a	a	a	a	a	a	a	a
Iceland	None	a	a	a	a	a	a	a	a	a	a	a
Ireland	None[22]	a	a	a	a	a	a	a	a	a	a	a

Table 7.A.1 School leadership appraisal practices (*continued*)

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
Israel	School principals; deputy school principal (all schools, ISCED level 1 and partially ISCED level 2 only)[23]	Mandatory periodic (every 3 years); in relation to a decision on employment status (contract renewal)	Central education authority or government	School principal: school inspector Deputy school principal: school principal	National standards for school leadership	Both school principals and deputy school principals: general leadership; pedagogical leadership; school climate; community relations; evaluation and accountability; interpersonal skills School principals only: organisation development; resource management	School principals: dialogue with evaluators; self-appraisal; school visit; school leader portfolio Deputy school principals: dialogue with the school principal; self-appraisal	Yes (5 levels)	Yes, it is expected to influence professional development activities	Yes, appraisal results influence decisions about promotion	None (under discussion)	Both school principals and deputy school principals: further appraisal; dismissal
Italy	None[24]	a	a	a	a	a	a	a	a	a	a	a
Korea	School Principal Appraisal for Professional Development (*): school principal; deputy school principal	Mandatory periodic (annually)	Central government	Regional education authorities (Metropolitan and Provincial Offices of Education); peer evaluators; teachers	A description of the general and professional duties of school leaders	General leadership; pedagogical leadership; organisational development; school climate; community relations; resource management; interpersonal skills	Teacher surveys; student surveys; parent surveys	Yes (5 levels)	Yes, it systematically results in a professional development plan	No	Extra professional development opportunities	Compulsory training
	Evaluation for School Management : School principal	At the discretion of regional education authorities (Metropolitan and Provincial Offices of Education)	Central government	Regional education authorities (Metropolitan and Provincial Offices of Education)	m	General leadership; parent and student satisfaction	Student outcomes; teacher surveys; student surveys; parent surveys	Yes (5 levels)	No	Yes, appraisal results may influence decisions about promotion	One-off financial bonus	Financial penalty
Luxembourg	None	a	a	a	a	a	a	a	a	a	a	a
Mexico	School principal (public central level schools at ISCED level 3 only)[25]	Mandatory periodic (annually)	Central education authority	Central education authority	National standards for school leadership	Organisation development; evaluation and accountability; resource management	Student outcomes	Yes (4 levels: excellent; good; adequate; insufficient)	Yes, it systematically results in a professional development plan	Yes, appraisal results influence decisions about promotion	None	Permanent contract not granted
Netherlands	None[26]	a	a	a	a	a	a	a	a	a	a	a

Table 7.A.1 School leadership appraisal practices (*continued*)

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
New Zealand	School principal	Mandatory periodic (annually)	Central education authority or government	School Board of Trustees (sometimes involving external appraisers)	National standards for school leadership	Achievement of a performance agreement in relation to general leadership; pedagogical leadership; school climate; community relations; and resource management.[27] Performance agreement includes one or more professional development objectives.	Indicators and information sources selected by school principals and evaluators; may include: dialogue with evaluators; self appraisal; student outcomes; student surveys; parent surveys; teacher surveys	No	Yes, it is expected to influence professional development activities	m	A staged career allowance	Further appraisal; professional development; statutory intervention
Norway	None[28]	a	a	a	a	a	a	a	a	a	a	a
Poland (*)	School principal (public schools only)	In relation to decision on employment status (contract renewal)[29]	Central education authority or government (general framework set by law)	Local education authorities; regional education authorities (education superintendents)	A description of the general and professional duties of teachers and school leaders defined in national regulations and legal requirements	General leadership; pedagogical leadership; organisation development; school climate; resource management	Outcomes of the external school evaluation; judgement of the school board (or pedagogical council of the school) and trade unions active in the school	Yes (3 levels: excellent, good, negative)	No	No	None	Contract not renewed
Portugal	School principal (public schools only)[30]	Mandatory periodic (every 4 years) There are 2 mandatory components: (1) internal appraisal; and (2) external appraisal	Central education authority	Internal appraisal (60% of final rating): General council External appraisal (40% of final rating): Inspectorate (results of external school evaluation)	Internal appraisal: mission statement (*carta de missão*) (prepared by school principal); leadership competencies as defined by school's general council; professional development requirements External appraisal: references used in external school evaluation by Inspectorate	Internal appraisal: Fulfilment of stated mission (*carta de missão*) (efficacy, efficiency, quality); leadership competencies; and professional development External appraisal: Aspects evaluated through external school evaluation (results, education service, leadership and management)	Internal appraisal: self-appraisal; appraisal of fulfilment of school principal mission statement (*carta de missão*); assessment of leadership competencies External appraisal: instruments used in external school evaluation by Inspectorate	Yes (5 levels) – there is a quota system for the top two levels at the national level, monitored by a national level Appraisal Co-ordination Council	No	Yes, appraisal results influence the speed at which a school leader progresses in the career structure	Extra service time to be credited on the next step of the career structure	Further appraisal; no service time counted; compulsory training
Slovak Republic	School principals; deputy school principals (all schools)[31]	In relation to employment status (contract renewal)	School organising bodies	School organising bodies	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body[32]	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body	Varies across schools depending on the school organising body
Slovenia (*)	School principal (public and government-dependent private schools only)	Mandatory periodic (annually)	Central education authority or government	School board	Rules on criteria for establishing work performance[33]	General leadership; pedagogical leadership; resource management	Dialogue with evaluators; school leader portfolio	Yes (rating on a scale 0–100)	No	No	One-off financial bonus	None

Table 7.A.1 School leadership appraisal practices (*continued*)

Country	Governance				Procedures				Use of results			
	For which school leaders is there a policy framework for individual appraisal?	Under which circumstances are school leaders appraised, and how often?	Who determines the procedures for school leader appraisal?	Who are the evaluators?	Against what references are school leaders appraised?	What aspects of school leadership are appraised?	What instruments and information sources are used?	Does the appraisal result in a rating for the school leader?	Does the appraisal inform the school leader's professional development activities?	Do appraisal results impact career advancement?	What rewards may school leadership appraisal involve?	What are the responses to underperformance of school leaders?
	1	2	3	4	5	6	7	8	9	10	11	12
Spain	School principal (public schools only)[34]	Mandatory periodic (annually)	State education authorities	State inspection bodies	A description of the general and professional duties of school leaders; school development plan or school project	General leadership; pedagogical leadership; organisation development; school climate; community relations; evaluation and accountability; resource management; interpersonal skills	Dialogue with evaluators; school leader portfolio (general annual report)	No	No	Appraisal influences the school principal's time in office and the consolidation of the salary level	Ongoing high performance through several positive appraisals allows school principals to remain at the salary level after leaving the school leadership position	Transfer
Sweden	None[35]	a	a	a	a	a	a	a	a	a	a	a
United Kingdom (Northern Ireland) (*)	School principal; deputy school principal (Performance Review and Staff Development Scheme [PRSD])	Mandatory periodic (annually)	Teachers' Negotiating Committee (Employing Authorities, Department and Teacher Unions)	Board of Governors (two reviewers); External Adviser	School Development Plan (SDP)	Three personal/shared objectives are set covering the areas of *(i)* leadership and management, *(ii)* pupil & curriculum development; and *(iii)* personal and professional development	*(i)* Classroom observations; *(ii)* task observations; *(iii)* review discussion	No, a review statement is prepared	Yes, the PRSD Scheme helps to identify the professional needs and necessary resources to support school leaders in their professional development	The PRSD Scheme helps to identify the professional needs and necessary resources to support school leaders in their career progression.	The PRSD Review Statement is part of the body of evidence used to inform decisions on pay progression.	There is an informal stage where a programme of support and development is provided. This may be followed by a formal stage which includes the issue of a formal written notice, a targeted support programme and ultimately dismissal if a satisfactory standard of work is not achieved.

Notes: a: information not applicable because the category does not apply; m: information not available; (*) The appraisal of school leaders is based on the teacher appraisal system.

1. Australia: Australia is a federation of eight states and territories. There are differences in school leader appraisal systems between states and territories and also between public (government) and private (non-government) schools. For most private schools, the school leader appraisal practices are set at the school level and carried out by the school board. Private schools that are part of a system (e.g. Catholic and Lutheran schools) also have frameworks for the appraisal of their school principals. While school principals' performance appraisal processes are undertaken by state-level jurisdictions, the National Professional Standard for Principals and the Australian Charter for Professional Learning of Teachers and School Leaders provide national guidance.

2. Austria: The policy framework for school evaluation considers school leadership as a general task, without a specific evaluation of individual school leaders. Only in case of a serious complaint, the state education authority can start an investigation on the allegations and appraise the performance of the school principal (rarely applied).

3. Belgium (Fl.): In public schools run by the Flemish Community of Belgium, school principals are also appraised at the completion of their probationary period of one year. If unsatisfactory, this probationary appraisal can lead to dismissal.

4. Belgium (Fl.): In public schools run by the Flemish Community, the role of the school organising body is delegated to the school group. In public schools run by provincial and local education authorities (provinces and municipalities), the local magistrate for education acts as the school principal's evaluator. In government-dependent private schools (including Catholic schools) evaluators mostly comprise volunteers.

5. Belgium (Fl.): School boards in both primary and secondary education are responsible for determining the responsibilities of the school principals The autonomy of the school board corresponds to the principle of Freedom of Education. The responsibilities for school principals are only pre-determined in public schools run by the Flemish Community.

6. Belgium (Fl.): In the case of underperformance, school principals without permanent appointment are dismissed from their function. If a school principal had previously been appointed to a permanent teaching position, they are allowed to take up their teaching responsibility.

7. Belgium (Fr.): Legislation explicitly refers to these three areas of competence of directors. Other aspects, however, can be appraised in addition to these three leadership domains at the discretion of the evaluators.

8. Belgium (Fr.): Legislation specifies that appraisal is a solely formative process.

9. Canada: Canada is a federation of ten provinces and three territories. There are differences in the appraisal of school leaders across provinces and territories.

10. Chile: In government-dependent private schools, school leaders are appraised at the discretion of their school organising body (sustainer).

11. Chile: The institutional targets need to be linked to a 2-4 year institutional objective. Among the institutional targets, at least two need to be linked to the School Management Quality Framework. In addition, at least one institutional target needs to be related to results, another to general leadership or pedagogical leadership. The institutional targets and professional development objectives are weighed at 50% each. For other teaching and technical-pedagogical leaders, the individually set institutional objectives need to be related to the institutional targets set as part of the school principal's appraisal.

12. Chile: In the case of a second unsatisfactory appraisal, the local education authority must inform the Municipal Council. The Municipal Council may remove a school leader from their functions with a two-third majority.

13. Chile: The appraisal result formed by the school organising body or school principal is subsequently verified by the provincial/regional education authority. In some cases, the result is also subject to an external auditing process.

14. Chile: At least one of the institutional targets needs to be related to results and a second institutional target to general leadership or pedagogical leadership.

15. Czech Republic: Following the introduction of new legislation, from 2012 onwards school principals in public schools are appointed for a period of 6 years. After this period, school organising bodies are required to appraise school principals to decide on the renewal of the contract. School organising bodies hold responsibility for determining the procedures for appraisal. The School Act (2004) serves as a reference standard for defining appraisal aspects. In public schools, school organising bodies also appraise school principals at their own discretion or in case of a complaint.

16. Denmark: ISCED levels 1 and 2 (public schools): Even though there is no central policy framework for appraisal, school leaders are appraised by local education authorities (municipalities). Practices vary between schools and municipalities. Some local education authorities (municipalities) require an annual appraisal of school leaders. Management by means of objectives is increasingly being used for school principals. Results contracts, school principal agreements and other forms of contracting serve as a means to define the objectives for the individual school (and school principal), typically for a 1- or 2-year period. Consequently, monitoring and performance systems are used to continuously assess if the school is performing according to the set objectives. Even though these instruments are implemented as management tools as such, they are equally important to hold schools and school leaders accountable for performance. There is no comprehensive overview of the instruments used to perform internal assessment of schools, but schools are likely to rely on various self-evaluation activities, which may involve a wide range of different methods of data collection. The external school assessment procedure is the annual quality report, which basically involves an external review by the municipal district council. ISCED level 3 (public schools): Annual result contracts are widely used.

17. Estonia: There is no central policy framework for the appraisal of school principals in place. The appraisal of school principals can take place at the local level, but is not mandatory.

18. Finland: There is no central policy framework for the appraisal of school principals in place. The appraisal of school principals is handled differently by each local education authority (municipality). In some cases, there is a personal results-based contract between the school principal and the local education authority (municipality), in which the outcomes and expected results of the school prinicpal's work are defined.

19. France: School principals in schools at ISCED level 1 are also often appraised as teachers. School principals in private schools are appraised by their employers. In addition, local inspection bodies (*Inspecteur de l'éducation nationale* [IEN]) assess the compliance of school principals in private schools with national education goals and programmes.

20. France: *Inspecteur de l'éducation nationale*.

21. Hungary: There is no central policy framework for the individual appraisal of school leaders in place. The appraisal of school principals takes place at the discretion of school organising bodies (maintainers). Central authorities merely recommend school organising bodies to appraise school leaders in the second and fourth year of their 5-year term. However, school organising bodies hold ultimate responsibility for determining if, when and how to appraise their school leaders. Appraisal typically involves an external expert. It typically takes the respective local quality management programme into account and school principals are often rated according to 4 grades (exceptionally suitable, suitable, less suitable, not suitable). Appraisal as carried out by school organising bodies typically fulfils both developmental and accountability functions. It may inform professional development activities and the re-appointment process. High performance may also be rewarded with an incremental or occasional salary bonus. Deputy school principals are appraised according to the local quality management programme of individual schools. The Hungarian appraisal system is envisaged to change substantially in 2013.

22. Ireland: Legislation in Ireland (Section 24 (3) of the Education Act 1998) allows for appraisal of a principal teacher as part of a process to address issues relating to professional competence or conduct.

23. Israel: A central policy framework for the appraisal of school principals in Israel was implemented in 2010-11 under the New Horizon reform programme. This framework applies to all schools at ISCED level 1 and partially to schools at ISCED level 2 except ultra-orthodox religious schools.

24. Italy: Italy has been developing a central policy framework for the mandatory periodic appraisal of school principals of public schools that appraises school principals every three years. It is envisaged that the evaluator team will comprise one inspector and two external evaluation experts.

25. Mexico: Information for the appraisal of school leaders in schools managed by state-level education authorities or autonomous agencies for all ISCED levels is not included in this table. At ISCED levels 1 and 2, the universal teacher appraisal system (see Table 5.1) and the National Teaching Career Programme (*Programa Nacional de Carrera Magisterial* [PNCM], see Table 5.4) also apply to school principals and pedagogical-technical advisors.

26. Netherlands: The appraisal of school leaders takes place at the discretion of school organising bodies (competent authorities). A central framework for the appraisal of school leaders has been in development, and is expected to be implemented in 2015. It is, however, not yet certain if appraisal and/or the central framework will be mandatory.

27. New Zealand: The professional standards for school leaders refer to school culture; pedagogy; systems; and partnerships and networks.

28. Norway: The appraisal of school leaders can take place at the local level.

29: Poland: Contracts for school principals contract are awarded for 5 years and directors are selected in a competitive process.

30. Portugal: Deputy school principals are appraised by the school principal as part of the teacher appraisal model.

31: Slovak Republic: There are national requirements for school leader professional development and the maximum length of contracts (5 years). School organising bodies (e.g. municipalities, regional education authorities, religious bodies) are legally required to conduct school leader appraisal, but are free to determine how they conduct this. Typically, appraisal involves an assessment of the achieved results in the school's development, a school leader's professional and leadership skills and participation in professional development. Evaluators often hold a dialogue with school leaders and draw on surveys, national measurements and school leader portfolios. Appraisal may also involve an analysis of school documentation.

32: Slovak Republic: Each employer (school organising body) creates its own ratings or may use the rating scale recommended by the Ministry (unsatisfactory, partially satisfactory, very good and exceptional).

33. Slovenia: Rules refer to the evaluation of work performance agreements covering leadership tasks according to a school principal's job description.

34. Spain: In the case of private institutions the employer is responsible for managing school leaders.

35: Sweden: Appraisal takes place at the discretion of local education authorities (municipalities, school boards).

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 7.A.2 Employment status and career development of school leaders

This table describes the employment status and career development of school leaders.

Country	Employment status			Career development	
	Who is the employer of school leaders?	What is the employment status of school leaders?	Can school leaders be employed on fixed-term contracts?	What is the structure of the school leader career?	What determines school leader career progression?
	1	2	3	4	5
Australia	State education authorities or governments; local education authorities; school, school board or committee[1]	Civil servant status; Salaried employee status	Yes, both school leaders with civil servant status and salaried employee status (maximum period of time varies nationally)	Multilevel career structure with a salary scale for each career level (number of levels varies nationally)	School leader appraisal results; merit selection process
Austria	ISCED level 1 (public schools): state education authorities ISCED level 2 (public schools): central or state education authorities depending on the type of schools ISCED level 3 (public schools): central education authorities	Civil servant status; salaried employee status	Yes, both school leaders with civil servant status and salaried employee status (for a maximum of 4 years)	Unique career stage with a single salary scale	Salary step increments based on length of service (biennial progression)
Belgium (Fl.)	School organising bodies (public schools and government-dependent private schools)	Civil servant status	Yes	Unique career stage with a single salary scale	Taking on extra roles and tasks
Belgium (Fr.)	School organising bodies (public schools and government-dependent private schools)	Civil servant status	No[2]	Unique career stage with a single salary scale	Length of service
Canada	School board (public schools)	Salaried employee status	Yes (maximum period of time varies across provinces/territories and is at the discretion of education authorities depending on the availability of permanent positions)	Multilevel career structure with a single salary scale (compensation tied to the size of the school)	Taking on extra roles and tasks; school leader appraisal results; exceeding expectations or exemplary performance in previous postings; supervisory officer qualifications programme; decision of head of school board
Chile	School organising bodies (sustainers)	Salaried employee status	Public schools: yes (5-year contract with possibility to re-apply) Government-dependent private schools: yes (for a maximum of 2 years)[3]	Public schools: unique career stage with a single salary scale Government-dependent private schools: at the discretion of school organising bodies (sustainers)	Public schools: salary step increments based on length of service Government-dependent schools: at the discretion of school organising bodies (sustainers)

Table 7.A.2 Employment status and career development of school leaders (*continued*)

Country	Employment status			Career development	
	Who is the employer of school leaders?	What is the employment status of school leaders?	Can school leaders be employed on fixed-term contracts?	What is the structure of the school leader career?	What determines school leader career progression?
	1	2	3	4	5
Czech Republic	School organising bodies (all schools)	Public schools: civil servant status Private schools: salaried employee status	Yes, school principals with civil servant status (6-year contract for all school leaders with the possibility of renewal); Yes, school principals with salaried employee status (at the discretion of school organising bodies)	All schools: unique career stage with a single salary scale	All schools: salary step increments based on length of service
Denmark	ISCED levels 1 and 2 (public schools): local education authorities ISCED level 3 (all schools): school boards	State civil servant status; salaried employee status	m	None	a
Estonia	Central government; local education authorities[4]	Salaried employee status	No	None	a
Finland	Local education authorities	Salaried employee status	No[5]	Unique career stage with a single salary scale (5 steps in schools with general programmes)	Certain steps on the salary scale according to the length of service
France	Central education authority (Ministry of Education)	Civil servant status	Yes	Multilevel career structure with a salary scale for each career level (3 career levels which respectively 10, 11 and 6 salary steps, each step corresponds to an indexed salary)	Appraisal results; length of service
Hungary	School organising bodies (maintainers)[6]	Public employee	Yes (5-year contract for all school leaders with the possibility of renewal)	Teacher multilevel career structure with a salary scale for each career level (5 career levels defined by qualification, 14 salary steps within each career level, progression on the salary scale every 3 years) plus school leadership bonus defined by the school organising body (maintainer) within a regulated percentage interval	Length of service; completion of professional development; taking on extra roles and tasks; school leader appraisal results
Iceland	ISCED levels 1 and 2: local education authorities ISCED level 3: central education authorities	ISCED levels 1 and 2: salaried employee status ISCED level 3: civil servant status	Yes, both school leaders with civil servant status and salaried employee status (for a maximum of 2 years)	ISCED levels 1 and 2: unique career stage with a single salary scale (18 steps) ISCED level 3: unique career stage with a single salary scale (9 steps)	ISCED levels 1 and 2: salary step increments based on length of service and age ISCED level 3: salary step increments based on length of service and school size

Table 7.A.2 Employment status and career development of school leaders (*continued*)

Country	Employment status			Career development	
	Who is the employer of school leaders?	What is the employment status of school leaders?	Can school leaders be employed on fixed-term contracts?	What is the structure of the school leader career?	What determines school leader career progression?
	1	2	3	4	5
Ireland	School Board of Management	Salaried employee status; state non-civil service status (public servant)	No	Unique career stage with a single salary scale (17 steps and allowance dependent on the school size)	Salary step increments based on length of service
Israel	ISCED levels 1 and 2: central government; corporation or non-profit organisation ISCED 3: local education authorities; corporation or non-profit organisation	ISCED levels 1, 2 and 3: civil servant status; salaried employee status	ISCED level 1 and 2: no ISCED level 3: yes, school leaders with salaried employee status only	ISCED level 1: multilevel career structure, with a salary scale for each career level (2 levels: 4 steps) ISCED level 2: multilevel career structure, with a salary scale for each career level (3 levels: 9 steps at level 2; 4 steps at levels 3 and 4) ISCED level 3: multilevel career structure within a single salary scale (3 levels)	ISCED levels 1 and 2: length of service; completion of professional development; school leader appraisal results ISCED level 3: length of service; completion of professional development
Italy	Central education authority (public schools only)	Civil servant status	No	Unique career stage with a single salary scale	Salary step increments based on length of service as a teacher; school context[7]
Korea	Central government	Civil servant status	No	Multilevel career structure with a single salary scale (2 levels, 50 steps)	Length of service; completion of professional development; appraisal results (Appraisal for School Management)
Luxembourg	Central education authority or government (public schools)	Civil servant status	No	Unique career structure with a single salary scale (salary scale defined for civil servants)	Salary step increments based on length of service
Mexico	ISCED levels 1 and 2 (public schools): state education authorities ISCED level 3 (public schools): central education authorities; state education authorities ISCED level 3 (all schools managed by autonomous agencies): autonomous and private institutions (e.g. universities)	ISCED levels 1, 2 and 3: salaried employee status	ISCED levels 1 and 2: no ISCED level 3: yes (varies between 6 months to 3 years)	ISCED levels 1 and 2: multilevel career structure, with a salary scale for each career level (through the National Teaching Career Programme [PNCM])[8] ISCED level 3: unique career stage with a single salary scale	ISCED levels 1 and 2: knowledge; aptitude; discipline; punctuality; length of service ISCED level 3: a; school leader appraisal results for public schools managed by central education authorities only
Netherlands	School organising bodies (competent authorities)	Public schools: civil servant status Private schools: salaried employee status[9]	Yes, both school leaders with civil servant status and salaried employee status (e.g. as interim director)	School leader career structure and number of salary steps can depend on school size	School leader appraisal results

Table 7.A.2 Employment status and career development of school leaders (*continued*)

Country	Employment status			Career development	
	Who is the employer of school leaders?	What is the employment status of school leaders?	Can school leaders be employed on fixed-term contracts?	What is the structure of the school leader career?	What determines school leader career progression?
	1	2	3	4	5
New Zealand	School Board of Trustees	Civil servant status; salaried employee status	Yes, both school leaders with civil servant status and salaried employee status	Unique career stage with a single salary scale	Salary step increments based on length of service and school leader appraisal results
Norway	Local education authorites[10]	Salaried employee status	Yes	Multilevel career structure with a salary scale for each career level (4 steps)[11]	School leader appraisal results; student results in some local education authorities
Poland	Local education authorities (communes, districts or other governing bodies)	Civil servant status; salaried employee status	Yes (5-year contract for all school principals with the possibility of renewal)	Unique career stage	a
Portugal	Central education authority	Civil servant status	No	Teacher career structure (unique career stage with 10 steps in the salary scale)	Length of service; appraisal results; completion of professional development (same as for teachers)
Slovak Republic	School and/or school organising bodies (in case of schools that do not have a legal personality)	Civil servant status	Yes (5-year contract for all school principals with the possibility of renewal)	None	a
Slovenia	School board	Civil servant status	Yes (5-year contract for all school principals with the possibility of renewal)	None	a
Spain	State education authorities (public schools); school organising bodies (government-dependent private schools)	State civil servant status	Yes (for a maximum of 4 years)	Unique career stage with a single salary scale (1 step)	a
Sweden	Local education authorities; school organising bodies (independent schools)	Salaried employee status[12]	Yes	Salaries are determined by local education authorities	a
United Kingdom (Northern Ireland)	Employing Authority[13]	Public servant; Salaried employee status	No	Single leadership pay scale for principals and vice-principals with placement linked to school unit total (enrolment)	School unit total (enrolment); school leader appraisal results

Notes: a: information not applicable because the category does not apply; m: information not available.

1. Australia is a federation of eight states/territories. There are differences in employment practices between states and territories, as well as differences between public (government) and private (non-government) institutions. In private schools that are part of a system (systemic non-government school systems), school leaders are often appointed by the local education authority. In independent private schools, school leaders are appointed by the school board or committee.

2. Belgium (Fr.): Except during the probationary period and if the school leader is substituting a school leader (e.g. on sick leave).

3. Chile: After two years, the contract automatically turns into a permanent type.

4. Estonia: Schools principals in public schools are appointed until their retirement. School principals in government-dependent private schools directors are appointed by school organising bodies. The salary of school principals is determined by the school owner (government, local authority or school organising body of a government-dependent private school).

5. Finland: School leaders are appointed until retirement.

6. Hungary: The trend towards a larger degree of central decision making in education in Hungary is also likely to influence employer responsibilities for school principals in 2013.

7. Italy: With the development of a central policy framework for the mandatory periodic appraisal of school leaders, it is intended that appraisal results are taken into account as one factor for determining progression on the salary scale. At present, school leaders' salaries are based upon the following two parts (according to the National Contract (CCNL) for school leaders): a) a fixed economic element; and b) a variable element determined at a regional level based on school leaders' responsibilities (85%) and outcomes (15%). The salary portion is attributed for responsibilities based on the following criteria: a) school dimension (number of students, number of teachers and other school personnel); b) complexity of the managed school (different school levels and school types within the same institution); c) geographical context (socially deprived areas, underdeveloped areas).

8. Mexico: *Programa Nacional de Carrera Magisterial.*

9. Netherlands: The terms of employment for school leaders as civil servants and salaried employees are identical.

10. Norway: According to the Management in Education Act, Section 9-1, each school is to have a sound professional, educational and administrative management. Teaching in schools shall be led by school principals. The school principals shall maintain familiarity with the day-to-day activities of the schools and endeavour to further develop the activities. Persons appointed as school principals must have pedagogical qualifications and the necessary leadership abilities.

11. Norway: The salary scale only regulates the minimum salary. The salary is for the most part locally determined and, in general, substantially higher than the minimum level.

12. Sweden: School leaders in Sami schools and special schools have civil servants status.

13. United Kingdom (Northern Ireland): The Employing Authority can be one of five Education and Library Boards, CCMS or Boards of Governors of Voluntary Grammar and Grant-Maintained Integrated Schools. With the implementation of the Education and Skills Authority (ESA) in 2013, ESA will become the employer for all principals in grant-aided schools.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Chapter 8

Education system evaluation:

Informing policies for system improvement

Within the evaluation and assessment framework, education system evaluation provides the opportunity to monitor quality and equity within the system and to bring together evidence from different aspects of the evaluation and assessment framework. This chapter presents evidence on different approaches to education system evaluation at both the national and sub-national levels. It examines governance issues, different procedures used, the capacity for undertaking and using the results of education system evaluation and the reporting of results. It then presents some options seeking to promote the better use of education system evaluation for informing policies for system improvement.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Introduction

This chapter looks at system evaluation within the evaluation and assessment framework. System evaluation refers to approaches to monitor and evaluate the performance of the education system as a whole, but also the performance of sub-national education systems, e.g. local authorities. The main aims of system evaluation are to provide accountability information to the public on how the education system is working and to inform policy planning to improve educational processes and outcomes.

System evaluation has a heightened role to play in the evaluation and assessment framework, as there is increased emphasis on evidence-based policy making. There is recognition that education is of central importance in shaping a knowledge society and increasing economic competitiveness.

The effective monitoring and evaluation of the education system is central to informing policy planning for improvement. In particular, system evaluation can provide valuable information to monitor equity within the education system and to help focus stakeholders on the major goals and challenges in the education system as a whole.

This chapter is organised in eight sections. After this introduction, the second section lays out the analytical approach, followed by a third section on impact, drivers and contextual developments. The following four sections describe key features of education system evaluation and country practices, structured along the four main topics of the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: governance, procedures, capacity and use of results. The final section provides pointers for policy development.

Analytical approach

Scope and definitions

The OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes conceptualises education system evaluation as the evaluation of an education system to provide accountability information to the public and to inform policies to improve educational processes and outcomes. The unit of evaluation can be either a national education system or a sub-national education system. Sub-national education systems are under the authority of a regional or local government or a group of private schools affiliated to a shared identity (e.g. pedagogical, religious, etc.).

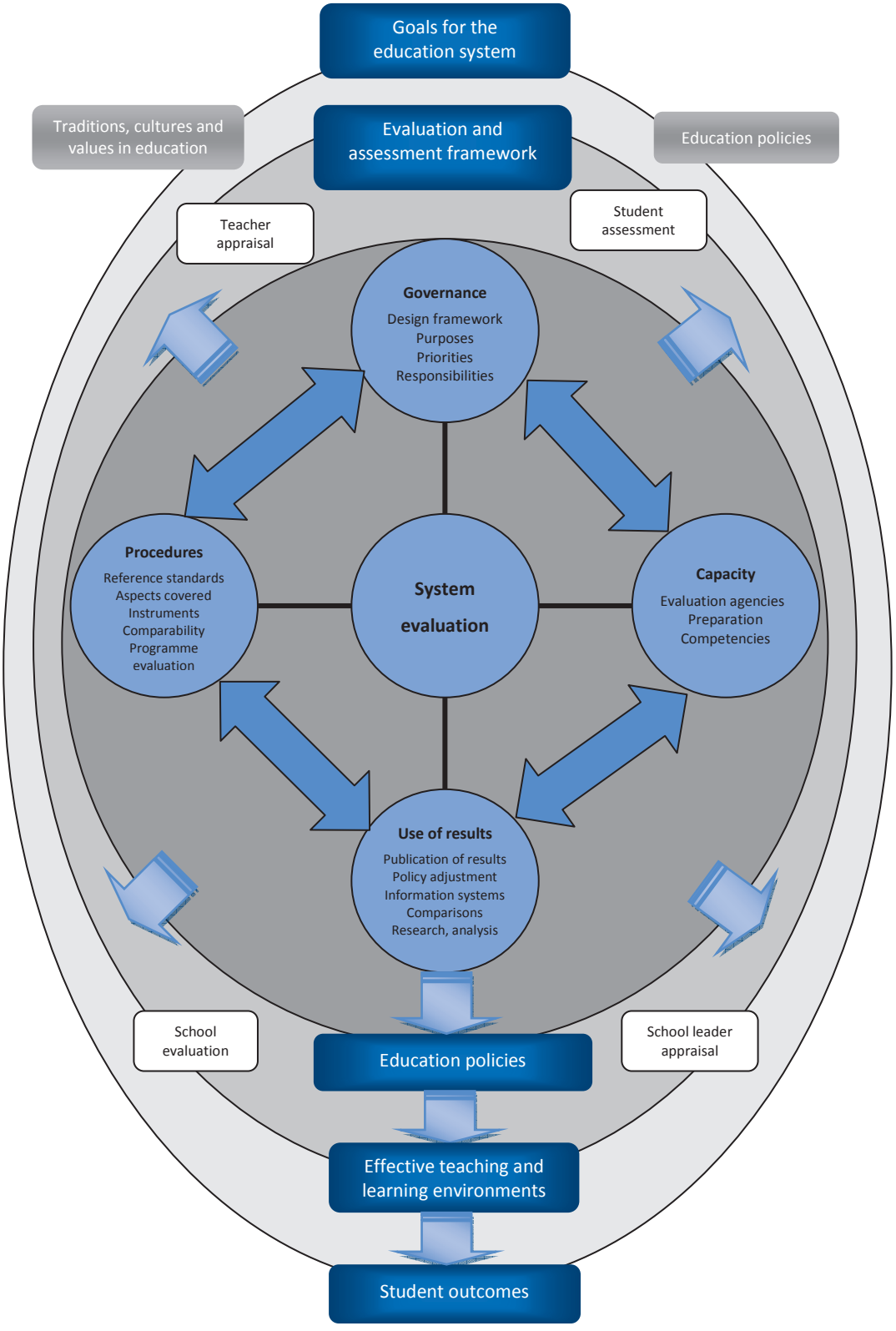
Different forms of evaluation may be used in education system evaluation. This chapter considers: the use of indicator frameworks to monitor key information on school systems; the use of tools to monitor student outcomes (in particular, specific national assessments designed for this purpose, longitudinal research and surveys, as well as international assessments); the use of qualitative reviews of particular aspects of the education system (including ad hoc reviews, as well as evaluative information generated via external education system reviews); and the evaluation of specific programmes and policies.

Conceptual framework

The OECD Review uses a conceptual framework to summarise the aspects involved in education system evaluation and the way these interconnect (see Figure 8.1). The overarching policy objective is to ensure that education system evaluation contributes to the improvement of student outcomes through improved education policies. There is a complex range of features associated with education system evaluation. This chapter presents these in four major areas:

- **Governing education system evaluation**: This addresses the purpose of education system evaluation and includes the major responsibilities for devising and conducting education system evaluation and for setting a legal framework. System evaluation concentrates on evaluating the achievement of student learning objectives at the system level, the achievement of education priorities including equity targets and assesses the teaching and learning environment. It also deals with the impact and implementation of specific policy initiatives. Finally, it refers to how education system evaluation is conceptualised in relation to the other components of the evaluation and assessment framework.
- **Procedures used in education system evaluation**: This aspect refers to the features of a given approach to system evaluation, that is, the mix of instruments, criteria and standards used in a specific system evaluation model. This may include the setting of specific targets and the development of different surveys to collect evidence on key features of the system and its performance.
- **Competencies to evaluate the education system and to use the results of system evaluation**: This aspect concerns the preparation to evaluate, to be evaluated and to use the results of an evaluation as well as the choice of the groups undertaking these functions. It includes issues such as: the choice of the evaluation agencies and the development of the skills to perform the evaluation of the education system; the preparation by education systems to be the subject of an evaluation; the development of competencies to effectively use the results of an evaluation for the improvement of education policies; and the design of agencies to review system evaluation results with a view to inform policy development.
- **Using the results of education system evaluation**: This encompasses the objectives of a system evaluation process and the mechanisms designed to ensure that evaluation results are used in a way such objectives are reached. The objectives of system evaluation typically consist of feedback for improvement of policies, accountability of policy makers and information about the performance of the education system and the impact of policies. Examples of mechanisms to use evaluation results include the systemic use of national monitoring results for school improvement, the development of information systems, the reporting on system performance including comparisons across schools and regions, policy adjustments, and research and analysis of system level data.

Figure 8.1 Conceptual framework for education system evaluation



Impact, drivers and contextual developments

The evaluation and assessment framework in OECD countries has been influenced by a number of contextual developments, notably decentralisation in many systems and the advent of new public management (see Chapter 2). At the same time, there is a general recognition of the need to use evidence to inform policies to improve educational processes and outcomes within the education system. There has also been growing demand across OECD countries to provide accountability information to the public on the performance of the education system. These general trends have increased the focus on education system evaluation within the evaluation and assessment framework. Monitoring systems have been developed to meet the demand for regular information on outcomes at different stages of the education system, typically via large-scale student assessments, but also via thematic evaluations of samples of schools as part of external school evaluation. In some systems, these have developed in parallel with the rise of school accountability systems. There is, therefore, a growing connection between school evaluation and system evaluation in several countries – in particular, in countries where sub-national education systems play a key role in school policy. This often reflects an approach to system and school evaluation as being part of a business planning cycle, in which priorities and targets are set for improvement based on an assessment of outcomes and processes in the system, and progress towards these targets is regularly monitored and evaluated. This is linked to the rise of performance measures and target setting in the wider public sector generally (see below).

Recognition of the importance of education in economic development

Arguably, a more specific influence on education system evaluation has been a renewed recognition of the importance of human capital in economic development and the changing demands for skills in the labour market. This has seen heightened prominence of education's role in – among other goals – preparing citizens to be productive members of the knowledge society. For example, in 1993, the Council of Ministers of Education, Canada (CMEC) stated:

> We are well aware of the challenges to the education systems posed by our rapidly changing world: globalization of the economy, openness with regard to other cultures, pressing needs for skilled labour, and technological advances that are having an impact on our daily lives as well as the job market. These changes require constant adjustments to our educational practices to ensure high quality, accessibility, mobility, and accountability.
>
> [*Joint Declaration: Future Directions for the Council of Ministers of Education,* Canada (CMEC), September 1993]

In this context, the profile of the results from international student assessments has been significantly raised in national policy discussions, with some politicians perceiving these as indicators of future economic competitiveness. For example, Australia and Denmark present examples of politicians setting aspirational targets for performance in OECD's Programme for International Student Assessment (PISA) (see Box 8.2). Major technological and methodological advances in international assessments have also raised the prominence of international comparative data in decision making (Tamassia and Adams, 2009).

A drive to measure outcomes for accountability in the public sector

There has been a general shift among OECD countries for public management to not only focus on financial inputs and processes (e.g. how inputs are transformed into goods and services), but also increasingly on outputs and outcomes. In general, public management literature distinguishes between outputs as the immediate result of government activity and outcomes as the final impact of this activity (OECD, 2009). For example, in relation to education, an output would be the number of children taught, but the outcomes would be what these children have learned. While good measures of both outputs and outcomes are vital, there are different measurement problems and interpretational concerns. An important distinction is that outcomes cannot be simply attributed to government actions or processes, as other factors outside the government's control are frequently involved. This has implications for the use of outcome measures in accountability systems and the assessment of performance against outcome targets can usually be done only generally.

Across OECD countries output/outcome measures have been introduced in budgetary procedures. The OECD (2007a) defines the accompanying set of performance measures as "the inputs, processes, outputs and outcomes used to assess the economy, efficiency and effectiveness of the activities of an organisation. They are quantitative or qualitative factors or variables that provide a means to measure achievement, to reflect the changes connected to an intervention, or to help assess performance." Across the OECD, government-wide measures on outputs and/or outcomes have been progressively introduced. Such measures typically include evaluation reports (e.g. programme, efficiency, sectoral or cost effectiveness reviews), performance measures, performance targets and benchmarks. The most commonly reported government-wide initiatives are evaluation reports and performance measures. In 2007, all OECD countries reported the existence of government-wide performance measures, except Belgium, the Czech Republic, Greece, Luxembourg, Portugal and the Slovak Republic. The majority of countries reporting the use of performance measures also reported the use of performance targets (the exceptions were Germany, Hungary, Iceland and Spain). Performance targets are considered to refer to specific outputs or outcomes that can be achieved in a shorter period of time than government goals or objectives (OECD, 2007a). Although performance targets typically are aligned against goals or objectives. Only ten OECD countries reported the use of government-wide benchmarking (Australia, Canada, Chile, Denmark, Finland, Germany, Mexico, the Netherlands, Norway and Sweden).

An increased role for system evaluation has also driven the demand for developing national assessments. The international achievement studies led by the International Association for the Study of Educational Achievement (IEA) raised the importance of educational assessment worldwide (Goldstein and Thomas, 2008). There was a need to contextualise the results from international studies in respect of the diversity in the goals and curricula used in different education systems. Also in both Denmark and the Netherlands, a strengthened focus on system evaluation has seen the introduction or strengthening of national monitoring systems to provide performance information on the system, in the context of a tradition of ad hoc evaluations of school development projects or policy programmes.

Already in 2003 a number of countries had introduced performance management at least partly to improve the accountability of agencies and ministries to the legislature and the public (OECD, 2004, p. 6). This drive to setting performance measures and targets in the public sector has led to the introduction of internal audit units within Ministries – and

this is common place in many countries. This has also promoted a culture of working with a view to showing demonstrable outcomes in end of year audit reporting. This has clearly influenced education system evaluation. For example, in the Flemish Community of Belgium, system evaluation is increasingly reliant on output measurement of predefined indicators (Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010, p. 57). In Australia, one of eight key strategies to support the achievement of the educational goals set for schooling across Australia as agreed by the different States and Territories in 2008 included "strengthening accountability and transparency" (DEEWR, 2010). Public reporting is an important part of system evaluation in Australia as it "enables the broader public to evaluate the system and the performance of governments" (DEEWR, 2010, p. 2).

Impact of system evaluation

The OECD Review has revealed a lack of research on the impact of system evaluation, per se. However, all countries in the Review have used the results of system evaluation at different stages, for example, to convene consultation on national curriculum or to refine or even establish education standards. It can, therefore, be seen that system evaluation strongly impacts policy and indeed serves informing policy as one of its major purposes (the other being to provide accountability information on the system).

Many systems have noted the shift to an increased focus on outcomes. Several have attributed this to a growing prominence of both national and international standardised assessment results in national policy and public debate. In countries such as Austria, Denmark and Luxembourg where there was no established tradition of monitoring outcomes of the education system, this shift is widely recognised as a result of the impact of international student assessments. For example, in Austria various studies have documented how the influence of international assessments (as part of system evaluation) has seen the introduction of education standards, the establishment of a specific agency to undertake evaluation and assessment and the development of national assessments. This pattern is mirrored in Denmark and Luxembourg. Several research studies have discussed the impact of international assessments on national policy making and the use of different results to argue positions by different stakeholders (e.g. Carvalho and Costa, 2009 for an overview; Rautalin and Alauutari, 2007; Grek, 2009; Delvaux and Mangez, 2008; Gür, Çelik and Özoglu, 2011). The PISA Governing Board has gained some further insight into this (Box 8.1).

However, there is quite some debate on the use of student test results in accountability systems. This is primarily based on research in the United States and typically related to the policy of No Child Left Behind. There is a contested debate and mixed evidence on the impact of test-based accountability systems (e.g. Rosenkvist, 2010; Morris, 2011; Masters, 2012). Many studies show evidence of improvement in test scores, but some studies demonstrate that this is due to artificial improvements or "test score inflation". There is far more consensus on the fact that test-based accountability systems exert a powerful influence over students' teaching and learning experience. Many studies demonstrate that this is often in a way that is not intended and may be detrimental to the actual quality of the students' learning experience.

Box 8.1 The impact of PISA on national policy making

In 2007, the PISA Governing Board commissioned a study to evaluate the impact of PISA results in participating countries (Hopkins et al., 2008). The study collected feedback via standardised questionnaires sent to policy makers, local government officials, school leaders, parents, researchers and media in 43 countries (548 questionnaires were returned), as well as to the PISA Governing Board members, and representatives from the business community and labour organisations. This was complemented by case studies in Canada, Hong Kong-China, Norway, Poland and Spain. Results showed that: PISA results were mainly used by policy makers, followed by local authority officials and school leaders; PISA is used to monitor and evaluate both performance and equity of the education system and has a high level of credibility and influence; countries and stakeholder groups increasingly value the skills assessed in PISA and promote these within their systems; the influence of PISA on both national and local policy formation is increasing, but has less impact on the school and classroom levels; PISA had made an impact on policy in all countries studied, but more so in countries with relatively low performance on the test, where many policy initiatives had been introduced directly as a consequence of PISA; the level of awareness among stakeholders varied across countries, but there was greater awareness in systems where policy makers and the media place more emphasis on results and generally in countries where average performance was lower. However, the study suggests that the media play the most important role in countries without comprehensive strategies for the dissemination of PISA results, which is likely to have a negative impact. PISA results were found to influence policies on curriculum revision, alignment of curriculum with assessment and instruction and accountability, e.g. with some countries introducing national testing and others refining their accountability frameworks. The study also identifies an emerging trend of countries aligning their assessment systems more closely with PISA. The study also identified some unexpected impacts including increased confidence in the education system, high levels of debate among stakeholders and a focus on regional differences and exploring reasons behind these.

In 2011, a short survey was administered to members of the PISA Governing Board to gain more insight into the impact of PISA and to explore to what extent PISA results were being used to evaluate and improve education system performance (Breakspear, 2012). Policy makers reported that PISA results were influencing policy to some degree in the majority of countries, regardless of their average performance on the test. The survey also revealed that PISA was being embedded in national policies to varying extent and in varying forms, via curriculum standards, assessment practices or performance targets. Indeed, the European Union has set specific target for its member countries linked to performance in PISA (see Box 8.3).

Sources: Hopkins et al. (2008); Breakspear (2012).

Governance

Purpose of education system evaluation

As with all components of the evaluation and assessment framework, education system evaluation serves both accountability and improvement. A major accountability objective is to provide information to the public on the general quality of the education system and feedback on reforms to the education system. The generation of such evaluative information should then inform policies to improve educational processes and outcomes. In general, six major aims can be distinguished: *(i)* to monitor student outcomes at a given point in time, including differences among different regions within the education system and given student groups (e.g. by gender, socio-economic or immigrant/cultural background); *(ii)* to monitor changes in student outcomes over time; *(iii)* to monitor the impact of given policy initiatives or educational programmes; *(iv)* to monitor demographic, administrative and contextual data which are useful to explain the outcomes of the education system; *(v)* to generate and feedback relevant information for

different agents in the education system; and *(vi)* to use the generated information for analysis, development and implementation of policies.

Responsibilities for education system evaluation

In the vast majority of OECD countries, the central education authority (e.g. Ministry of Education, Department of Education) is responsible for the development of education system evaluation (see Table 8.1). However, in Australia, Canada, the Czech Republic, the Slovak Republic, Spain and Sweden responsibility is shared between central and decentralised education authorities. In a number of Canadian provinces/territories, system evaluation is very much aligned with school evaluation, as it is conceived as "a business model" in which education departments at the provincial/territorial level, school boards or districts and schools are mandated every year to review and refine their strategic goals (Fournier and Mildon, forthcoming). However, in reality other bodies also play a role in education system evaluation, notably: the Council of Ministers of Education, Canada (CMEC) and regional bodies such as the Council of Atlantic Ministers of Education and Training (CAMET/CAMEF). In Austria, Chile, Denmark, Estonia and New Zealand, the central education authority shares this responsibility with specific evaluation bodies. In Italy, a specific central agency is responsible for the development of education system evaluation. Indeed, most countries draw on the capacity of a range of entities in conducting education system evaluation, including quality assurance agencies, inspectorates or school review agencies, audit offices and education councils bringing together a range of stakeholders (see below).

A framework for education system evaluation

The OECD Review has revealed that in many systems there is not an overall framework for education system evaluation, but rather a suite of different elements in the legal framework for governing schooling that has gradually been built up and strengthened the role of different types of evaluation within the system over the years. Indeed, Denmark, Korea, Norway and Sweden do not have a policy framework in place for the evaluation of the education system (see Table 8.1). In Hungary and Portugal, the legal framework for education system evaluation is included in basic laws for education (note that Hungary is introducing substantial changes to the basic law in 2013). For example, in Portugal: "the education system must be the object of continuous evaluation and must take account of the educational and pedagogical, psychological and sociological, organisational, economic and financial, and politico-administrative and cultural aspects" (Portuguese Ministry of Education and Science, forthcoming). Many systems with external school evaluation mechanisms in place (e.g. school inspectorates or school review bodies) have policy frameworks for the evaluation of the education system and in general this applies to all schools within the system. In such systems, there is a specific legal framework in place governing the external evaluation of schools and evidence from such external reviews can feed into education system evaluation (see Chapter 6). There may also be specific laws pertaining to the assessment of students within the system and again results from such assessments may feed into education system evaluation. Other laws may be more specific to the collection of information for education system evaluation. For example, in Austria, the Austrian Education Documentation Act 2003 (amended in 2008) introduced the possibility to monitor educational developments longitudinally.

Table 8.1 Policy frameworks and development of education system evaluation (2012)

	National policy framework to evaluate education system	Bodies responsible for developing education system evaluation
Australia	All schools; also state-level frameworks for public schools	Central education authority; state education authorities
Austria	Public and government-dependent private schools	Central education authority; central agency (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[1]
Belgium (Fl.)	All schools	Central education authority
Belgium (Fr.)	All schools	Central education authority
Canada	Public schools	Provincial/territorial education authorities
Chile	All schools	Central education authority; central agency (The Quality of Education Agency)
Czech Republic	All schools	Central education authority (Ministry of Education); Czech School Inspectorate (CSI); regional education authorities
Denmark	No framework	Central education authority; central agency (Quality and Supervision Agency)
Estonia	All schools	Central education authority (Ministry of Education and Research); central agency (Foundation Innove)
Finland	All schools	Central education authority (Ministry of Education and Culture)
France	Public and government-dependent private schools	Legislative through laws; executive through decrees and circulars
Hungary	The Law on Education has this role	Central education authority (The Educational Authority)
Iceland	All schools	Central education authority
Ireland	Public and government-dependent private schools	Central education authority (Department of Education and Skills [DES]; Inspectorate of the DES)
Israel	All schools	Central education authority
Italy	All schools	Central agency (National Institute for the Evaluation of the Education System [INVALSI])
Korea	No framework	Central education authority
Luxembourg	No framework	Central education authority
Mexico	All schools	Central education authority
Netherlands	All schools	Central education authority (Dutch Inspectorate of Education)
New Zealand	Public schools	Central education authority; central agency (Education Review Office)
Norway	No framework	Central education authority (Ministry of Education and Research; Norwegian Directorate for Education and Training)
Poland	All schools	Central education authority (Ministry of Education)
Portugal	All schools	Central education authority (General Directorate for Education and Science Statistics; General Inspectorate of Education and Science)
Slovak Republic	All schools	Central education authority (Ministry of Education, Science, Research and Sport); State School Inspectorate (SSI); regional education authorities
Slovenia	All schools	Central education authority
Spain	All schools	Central education authority; state education authorities
Sweden	No framework	Central education authority; local education authorities
United Kingdom (Northern Ireland)	No framework	Central education authority (Department of Education)

Many countries lack an overall strategic approach to education system evaluation. This is reflected, for example, in how countries plan and organise the collection of evidence for education system evaluation. It is not established practice among OECD countries to map out the available information for education system evaluation against education system objectives and to draw up a plan for the further collection of information. Among systems

participating in the OECD Review, this is the case only in Australia, the Czech Republic, Hungary, Israel, the Netherlands and the Slovak Republic. A total of ten systems map existing information against education system priorities (Table 8.2).

Table 8.2 Indicators of a strategic approach to education system evaluation in OECD countries (2012)

A plan to prioritise further collection of information and a mapping of existing information against education system priorities	Australia; Czech Republic; Hungary; Israel; Netherlands; Slovak Republic
A mapping of existing information against education system priorities	France; Iceland; Ireland; Northern Ireland (UK)
A plan to prioritise further collection of information	Belgium (French and Flemish Communities); Chile; Finland; Slovenia; Spain
Neither	Austria; Denmark; Italy; Korea; Luxembourg; Mexico; New Zealand; Norway; Poland; Sweden

Note: Canada – all provinces/territories either have a mapping in place or plan the prioritisation of information collection.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Placing the framework for education system evaluation in the broader context of public services evaluation

Increasingly public services evaluation is driven by performance management which may present challenges to align education system evaluation practices. This may create demand for the collection of new evidence and measures on outcomes. In some systems, there may be a lack of comparable measures of student outcomes and/or only limited measures available for particular stages of education and/or in discrete skills. Some systems have noted an over dependence in national education policy development on the results of international measures of education system outcomes (e.g. OECD and IEA international student assessments) in the absence of national measures.

Further, in arguably the majority of systems, there is only an emerging culture of systematically evaluating the impact and outcomes of different educational interventions and again these efforts may be hindered by a lack of reliable and comparable information on student outcomes. In the Netherlands, programme evaluation had been the major form of system evaluation in the 1970s. Evaluation committees comprising educational experts had evaluated various programmes, but had to depend on "fragmented sets of coincidental information". None of the programme evaluations were designed to identify the link between the policy programme and outcomes. However, more recent programme evaluations can draw on information available from the national monitoring system. The programme evaluation approach was gradually replaced by a monitoring approach with the introduction of new evaluation instruments including a national sample assessment in primary education and cohort studies in primary and lower secondary education. This was coupled with "a strong urge for economisation and budget control" (p. xxiv, Scheerens et al., 2012).

In Sweden, since the early 1990s the principle of governing through goals and objectives combined with national standard-setting has been highly applied in the education sector (Nusche et al., 2011a). Two external bodies may monitor the education sector: the Swedish National Audit Office and the Swedish Agency for Public Management (Nusche et al., 2011a). The Swedish National Audit Office not only produces 30 reports each year on the whole of the public sector (financial audit) but also audits effectiveness in different areas, including education. The Swedish Agency for Public Management is under the Ministry of Finance and is responsible for conducting quality surveys/evaluations. In Denmark, the Audit of the State Accounts is trying to promote the need for Ministries to

conduct studies on effects and outcomes (Shewbridge et al., 2011). It finds that the principles of performance management are not yet firmly embedded in the Ministry of Education. The expectation is for the Ministry of Education to conduct evaluation studies on the effectiveness of its policies and that national audits remain limited to checking on the correct implementation of government policies and not on their effectiveness. In Australia, the Australian Government Productivity Commission is an independent statutory authority with a major role in monitoring education outcomes in Australia (Santiago et al., 2011). This authority enjoys a fair degree of freedom, although the Government commissions enquiries on a range of economic, social and environmental issues. The Productivity Commission monitors national education and other government sectors on a set of agreed indicators in the annual Report on Government Services (see Box 8.2).

Box 8.2 The influence of developments in public sector performance management

Calls to monitor public sector performance

In Australia, there was a shift in focus in general government reporting from inputs to results in 1999-2000 and the first Budget report on an accrual-based outcomes and outputs framework (Australian National Audit Office/CPA Australia, 2008). In 1997 a national taskforce was established to oversee development of literacy and numeracy benchmarks for students in Years 3, 5 and 7. This led to the publication for the first time in the 1999 National Report on Schooling in Australia of nationally comparable data on student performance against the Years 3 and 5 reading benchmarks. In subsequent issues of the report, nationally comparable data on literacy (reading and writing) and numeracy for Years 3, 5, 7 and 9 were published as they became available. In 1999, another key national taskforce was responsible for developing further performance measures for national reporting against national goals. Participation and attainment measures were then agreed and first reported in the 2000 National Report for Schooling. The Report on Government Services' Performance Indicator Framework provides a common reporting basis for the Australian Government and each state and territory government. A recent independent review of the framework highlighted the potential efficiency of sharing information on performance indicator methodology as well as some performance measures across government services (Steering Committee for Review of Government Service Provision, 2010).

In New Zealand, new expectations were introduced in 2001 for public service departments to adopt a more strategic and outcomes-focused approach to management and reporting. The underlying rationale for the Managing for Outcomes initiative is that departments will be able to demonstrate how activities contribute to desired results or outcomes and will be able to make more informed decisions on future interventions. This encompasses the following management cycle: setting direction – what do we intend to achieve over the next three to five years and why?; planning – what is the best way to achieve this and have we got the required capability?; implementing and delivering – are we implementing and delivering as planned, and managing our capability and risks effectively? This implies the monitoring of interventions; review results – what impact have our interventions had and what improvements can we make? This implies the assessment of the effectiveness and efficiency of interventions and considering unintended consequences. Government departments are expected to develop a Statement of Intent for a 3-to-5-year period and to produce annual reports on progress against this. For the period 2009-14, the Ministry of Education specified six priority areas "to ensure a clear focus on raising standards across the education system" (Ministry of Education, 2009). For each area the Ministry of Education details what it is seeking to achieve; how it will demonstrate success (with specified performance indicators); what it will do to achieve this (including specific measures to judge Ministry performance). Further, these priority areas guide the work of the Ministry of Education and the educational agencies, e.g. the Education Review Office and the New Zealand Qualifications Authority.

Box 8.2 The influence of developments in public sector performance management (*continued*)

The economic discourse

In Denmark, in 2006 the Government established a strategy for Denmark in the global economy in which it is stated that Danish students shall be among the best in the world in reading, mathematics, science and English – this was reinforced by a specific target set in 2010 for Danish students to be in the top five countries as judged in international assessments (Regeringen, 2010). Since 2006, an annual "Competitiveness Report" is published which includes indicators on the education system. These wider reform processes (e.g. work by the globalisation council and the government growth forum) have driven the production of performance data on schooling.

In Australia, at the highest political level there has been recognition of the importance of securing high-quality educational opportunities and outcomes for Australian students. As in many other OECD countries, politicians cite the importance of education's role in securing the nation's future productivity and international competitiveness. As such, education has a prominent place in the 2008 Council of Australian Governments (COAG) National Productivity Agenda for reform. COAG has set three major targets for schooling including an increased proportion of young Australians attaining senior secondary education and two targets to reduce the performance gap of Indigenous students. The National Plan for School Improvement, announced in September 2012, includes a new target to place Australia among the top five schooling systems in the world by 2025 in mathematics, science and reading achievement and for providing a high quality, high equity education system. This target is reiterated in the Australian Government's Australia in the Asian Century White Paper, released in November 2012.

Ensuring a comprehensive approach to system evaluation

In many countries, different elements used for education system evaluation have been established at different times and may evolve to adapt to different needs. The challenge for many systems is to design a comprehensive approach to education system evaluation which encompasses these different purposes. Further, in systems where the major evaluation responsibilities lie at the sub-national level, there may not be a comprehensive overview of evaluation capacity at the sub-national level. Varied quality and coverage of sub-national monitoring systems may pose significant challenges to ensuring equitable schooling opportunities and outcomes for students nationally.

Procedures

References and standards used in system evaluation

The OECD Review has revealed that countries use a variety of references and standards in evaluating education systems. In the broadest sense, education system evaluation aims to evaluate the extent to which the education system has met national goals and objectives for the education system. These can range, for example, from specific goals to provide high-quality education to students, to promoting national values and civic responsibilities and providing highly skilled individuals to generate economic productivity (see also Chapter 3). There may be particular attention given to equity within educational provision and the need to improve educational outcomes for particular student groups. Such national goals may also be complemented by specific goals and objectives set at the local level. For example, some sub-national systems may face specific economic or social challenges and reflect these in goals for the sub-national

education system. The local economy may demand specific skills or may face particular employment needs or concerns in different sectors. Further, additional goals for sub-national systems could reflect a specific pedagogical approach, ethos or religious values.

However, it is a typical approach across countries to use a mix of different references and standards for specific forms of education system evaluation. For example, in Luxembourg the overall performance of the education system is evaluated "by putting together and matching different outcomes from several bodies" (ADQS, 2011). In the Slovak Republic, education system evaluation draws from many different elements that may not be "mutually compatible" (Hajdúková et al., forthcoming). In Portugal there is "no explicit overall device" for education system evaluation, but results from different structured sectoral instruments feed into the evaluation of different aspects of the education system (Portuguese Ministry of Education and Science, forthcoming).

One major instrument used by countries is national assessments and these typically aim to monitor the implementation of national curriculum and/or student progress against specific student learning objectives or educational standards (see also Chapter 4):

- National curriculum goals: Chile (ISCED 3), Estonia, Finland, Iceland, Ireland, Israel, Italy, Korea, Mexico, New Zealand, Portugal, Slovak Republic, Slovenia and Sweden;
- National curriculum goals and standards: Chile (ISCED 1 and 2), Czech Republic, France, Poland and Spain (currently being developed);
- National standards: Austria, Belgium (the Flemish and French Communities) and Luxembourg;
- National learning progressions: Australia, Denmark, the Netherlands, Norway and Northern Ireland in the United Kingdom.

In other systems, the specific assessment framework will be the major reference for national assessments, for example, this is the case for the Pan-Canadian Assessment used in Canada and for the National Assessment of Basic Competencies in Hungary.

The extent to which countries set specific performance expectations of student learning in different educational areas varies significantly (see Chapter 4). For example, in Austria, there are currently no definite performance criteria for system evaluation, but educational standards based on performance norms in the new national assessments will be established (Specht and Sobanski, 2012). The first national assessments are being conducted in 2012. In France and Luxembourg, the expected core competencies (and expected knowledge in France) are used as standards to monitor student progress and define both the expected minimum level of competencies to be acquired by students, as well as the acquisition of higher levels of competencies.

A second set of reference points for education system evaluation can be observed in the increased use of comparative measures to judge improvements or decline in specific assessments. For example, by comparing empirical data in cross-sectional cyclical international studies and also in longitudinal national studies to monitor decline or improvement over time. At the highest political levels, there may also be examples of setting specific goals to improve country average performance relative to other countries in international assessments.

A third set of reference points for education system evaluation may be objectives set for specific education policies. The evaluation would comprise a judgement of how well

these objectives have been achieved. This is the major approach used in Korea, where there is an evaluation of major policies and systems approximately every ten years, e.g. the Ministry of Education's Forty Years of Education published in 1988 and the government's special planning team's Forty Years of Korean Education in 2007. Plus, at the start of each new government, there is an evaluation of the previous administration's education policies. Further, there is an evaluation of the major education policies around the middle to end of a presidential term (Kim et al., 2010).

Setting targets for system performance

In several systems, major references for education system evaluation are the policy priorities of governments in office and related education targets (see Chapter 3). Countries may set specific targets set to be achieved over a certain timeline and education system evaluation may monitor progress towards achieving such targets. For example, in the Netherlands, current policy agendas are increasingly expressed in terms of explicit attainment targets. This leads to a more straightforward interpretation of information from education system evaluation procedures (Scheerens et al., 2012). In Australia, national targets related to schooling have been agreed and align with performance measures outlined in the *Measurement Framework for Australian Schooling* (ACARA, 2011). This framework provides the basis for government reporting on the performance of schooling in Australia. A national target is defined as "a measurable level of performance expected to be attained within a specified time" (idem).

In Mexico, education system evaluation was framed by the priorities set in the 2007-12 five-year National Development Plan and an Education Sector Programme. The Education Sector Programme set six clear policy objectives for the education sector and specified a set of 41 indicators (22 of which were in basic education). Each indicator included a target to be achieved by 2012 and how this would be measured. For example, one of the six objectives was the promotion of ICT in education. The 2012 goals refer to equipping media rooms, increasing the number of computers per student, establishing Internet connections in libraries and training teachers in the educational use of computers and ICT. The establishment of clear policy objectives along with indicators and targets helps provide a reference in relation to which the relevance and effectiveness of education policies can be measured. It also ensures greater focus on the main challenges the education system is facing and encourages stakeholders at all levels to develop strategies responding to these (Santiago et al., 2012).

In Northern Ireland in the United Kingdom, target setting is a key part of education system evaluation (Department of Education, forthcoming). Targets are set towards achieving the Minister's long-term goals for 2020 and progress towards these is monitored. Targets are set at the highest level in the Programme for Government 2011-2015 which includes high-level targets for the performance of the education system by 2015, with interim milestone targets. For example, a target is set for 2014/15 that 66% of young people achieve at least five General Certificates in Secondary Education with a mark of A to C in mathematics, English and three other subjects. The interim targets are 61% in 2012/13 and 63% in 2013/14. The setting of targets allows policies to be adjusted based on robust evidence of the rate of progress. For example, the Minister's long-term targets were in fact reviewed and increased recently based on an assessment of the rate of progress.

Targets for the education system may also be set at the supra-national level (see Box 8.3). In the Flemish Community of Belgium, the publication of European Union (EU) benchmarking reports is perceived as an indirect evaluation of the Flemish

education system (Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010, p. 48). It is of note, also, that the EU has included a target specifically linked to results in the OECD's PISA survey. This inevitably raises the profile of PISA results in member countries.

Box 8.3 Benchmarks for education systems: European Union

The European Union sets objectives for its member education systems. The following benchmarks are supposed to be achieved by 2020:

- At least 95% of children between the age of four and the age for starting compulsory primary education should participate in early childhood education.
- The share of 15-year-olds with insufficient abilities in reading, mathematics and science should be less than 15% (specifically as measured by and defined in OECD's PISA).
- The share of early leavers from education and training should be less than 10%.
- The share of 30-34 year-olds with tertiary educational attainment should be at least 40%.
- An average of at least 15% of adults (age group 25-64) should participate in lifelong learning.

This has prompted strategies to monitor results in member countries and the setting of specific targets within countries. For example, Portugal has developed the following annual indicators to monitor progress towards the achievement of the targets: *(i)* early leaving rates at the ages of 14, 15 and 16; *(ii)* the proportion of students who repeat grades in the first, second and third cycles of basic education and in secondary education; *(iii)* the proportion of students with a mark above the passing level in each national Portuguese language and mathematics tests/examinations in both basic and secondary education.

Sources: European Commission, Lifelong Learning Policy, *http://ec.europa.eu/education/lifelong-learning-policy/framework_en.htm*; Santiago et al. (2012).

Methods and instruments to assess the performance of the education system

A range of tools can be used to measure the performance of the education system. This section presents an overview of the major tools used in OECD countries.

Indicator frameworks

A major driver for the collection of information on the education system has been the joint international standardised data collection by UNESCO, OECD and EUROSTAT. The international education indicators framework covers, as main areas, participation (student enrolment in compulsory schooling and entrance to tertiary institutions), outcomes (educational attainment of the adult population; student outcomes including graduation and dropout rates), teaching workforce (demographics; qualifications; salaries and working conditions) and finance (annual expenditure by educational level and school type). Data collection procedures are well established and have drawn on and influenced national indicator frameworks.

At the national level, indicators are often reported in an annual publication with statistics and indicators on education. For example, this is the case in 20 of the systems participating in the OECD Review (Table 8.9). Many systems have developed national

indicator frameworks (see for example Box 8.4). These allow the monitoring of the education system over time against a stable set of measures. Such frameworks can be augmented over time with new measures according to national priorities and may also be complemented by specific indicator sets established to monitor priority areas.

Box 8.4 The development of indicator frameworks for system evaluation in Australia and Canada

Canada

As early as 1994, the regional institution Canadian Atlantic Ministers of Education and Training (CAMET/CAMEF) suggested the development of a set of education indicators within the context of consultations in the pan-Canadian organisation, the Council of Ministers of Education, Canada (CMEC). The Pan-Canadian Education Indicators Program has the major objectives "to develop and maintain a set of statistics that provide information about education and learning in Canada and to support evidence-based policy making" (Canadian Statistics Education Council, 2012). Since 1996, a regular set of indicators has been published. In 2010, in a joint declaration of provincial and territorial ministers of education, Learn Canada 2020, underlined the importance of system evaluation by setting eight specific objectives for "elementary and high-school systems", including:

- **Learning Assessment Programs and Performance Indicators**: Support the implementation of national and international learning assessment programs and performance indicators for education systems.
- **Education Data and Research Strategy**: Create comprehensive, long-term strategies to collect, analyse, and disseminate nationally and internationally comparable data and research.

In response, the Canadian Education Statistics Council has published a framework for the collection of indicators on learning and education in Canada whose goal is to provide information for a variety of purposes, including (CMEC, 2010): to describe and compare educational systems, programs, learners, etc. at a pan-Canadian level and internationally with other OECD countries; to assist policy development; to direct program administration; to monitor and evaluate effectiveness and efficiency; to engender research on educational issues; to inform learners (and their families) about their learning options; to provide evidence on accountability on the effective use of public funds. The indicators framework aids the process of deciding information priorities for system evaluation, ensuring that important information gaps are addressed and unnecessary overlap and duplication is avoided, as well as identifying the relevance of the information to the system and how information is interrelated.

Australia

A core strength of system evaluation in Australia is the existence of clear standard frameworks both for reporting key performance measures and for general government sector reporting (see Box 8.2). Since 2000 all Australian Education Ministers have worked on producing a common measurement framework including national Key Performance Measures. Ministers first defined national Key Performance Measures in early 2000 as "a set of measures limited in number and strategic in orientation, which provides nationally comparable data on aspects of performance critical to monitoring progress against the National Goals for Schooling in the 21st Century" (MCEETYA, 2008). This framework clearly presents the agreed measures and their source for each of the priority areas: literacy, numeracy, science literacy, civics and citizenship, information and communication technologies (ICT) literacy, vocational education and training (VET) in schools, student participation, student attainment and student attendance. In 2008, the framework was enhanced by the inclusion of comparable measures from the National Assessment Program – Literacy and Numeracy (a full-cohort assessment). The framework was reviewed and in late 2010 was further refined to incorporate the full suite of agreed national key performance measures, including those addressing the key targets set by the Council of Australian Governments in its National Productivity Agenda (see Box 8.2). The *Measurement Framework for Schooling in Australia 2010* defines the national key performance measures, specifies data sources for these measures and outlines the reporting schedule for 2010-2015. This measurement framework was most recently reviewed in 2012 and will continue to be reviewed every three years by the Australian Curriculum, Assessment and Reporting Authority.

Sources: CMEC (2010), *http://cmec.ca/Publications/Lists/Publications/Attachments/257/cesc-data-framework-sept2010.pdf*; ACARA (2011), *www.acara.edu.au/verve/_resources/Measurement_Framework_for_Schooling_in_Australia_2010.pdf*.

In New Zealand, an Education Indicators Framework has been developed to help decision makers analyse the state of the education system and monitor trends over time (Nusche et al., 2012). The indicators described in this framework relate to six priority areas: education and learning; effective teaching, student participation; family and community; quality education providers; and resourcing. For each of these six indicator domains, there are specific measures to determine the extent to which certain aspects of a result have been achieved. The Indicator Framework also includes contextual information to help the interpretation of results.

In Norway, the Directorate for Education and Training uses a stable and common reporting framework to evaluate the Norwegian school system (Nusche et al., 2011b). The framework includes five core areas: learning outcomes; learning environment; completion rates in upper secondary education; resources; and school facts. This systematic approach clarifies the national reporting process and ensures harmony across different reporting mechanisms (see Box 8.11). Norway also makes use of periodic monitoring of priority policy areas. A recent example is a specific indicator system to monitor quality in teacher education and in the profession (GNIST). The monitoring system was implemented in 2008 and contains five target areas (recruitment, quality in education, quality in teaching, quality in school leadership, improved status for the profession) with 23 indicators to monitor improvement/progression. The basic approach is to make use of existing information available nationally, but to highlight this in a coherent set of indicators. At the same time, GNIST has used some firsthand research, e.g. via the administration of surveys to teacher educators, school principals and teachers on their perception of quality in education.

In France, there are dual indicator frameworks. One framework aims to provide accountability information on the performance of the education system in line with wider government performance monitoring. This includes indicators on: effectiveness, e.g. the rate of grade repetition; efficiency, e.g. efficiency ratio for human resource management; and financial indicators, e.g. operating expenditures. A second indicator framework is used for the three major statistical publications on the education system produced by the DEPP: one presents 30 indicators on results of the education system and trends over time; the second one focuses on different regions with indicators on economic and social environment, school context, human resources, student pathways and results; and the third one is composed of all available statistical information on the educational system's functioning and results.

National assessments

In all OECD countries except Greece and Portugal (and Scotland within the United Kingdom), national assessments are administered during primary and/or secondary schooling. The OECD (2011) defines national assessments as "similar to national examinations in that they aim to measure the extent to which students have acquired a certain amount of knowledge in a given subject. National assessments may be mandatory but they do not have an impact on students' progression or certification as examinations do. Assessments are mostly used to monitor the quality of education at the system and/or school level. They also provide feedback to improve instruction and show the relative performance of students." These are administered in two major forms: full cohort, i.e. each student in the given school year is tested (with exemptions only for certain students as defined nationally, typically those with severe cognitive disabilities, but also students in isolated communities); sample, i.e. the assessment is administered in a selection of schools with students in the given school year (typically, the choice of

schools aims to give a nationally representative sample and pays attention to a balanced mix of types of school, socio-economic composition of students in school, school location, or any other relevant factors that are highly associated with student achievement nationally; students may also be randomly sampled within schools). Table 8.3 presents an overview of full-cohort and sample national assessments in OECD countries.

Table 8.3 National assessments in OECD countries (2012)

	National assessments (excluding international student assessments)
Australia	FULL COHORT – Years 3, 5, 7 and 9: reading, writing, language conventions (spelling, grammar and punctuation) and numeracy SAMPLE – Years 6 and 10: ICT, civics and citizenship; Year 6: science.
Austria	FULL COHORT – Years 4 and 8: mathematics and German; Year 8: English.
Belgium (Fr.)	FULL COHORT – ISCED 1, 2 and 3: subjects are rotated on the basis of 3-year cycle, e.g. mathematics in 2011/12. A 3-year testing plan is developed for this. The test is conducted for the purpose of identifying individual learning needs.
Belgium (Fl.)	SAMPLE – No systematic assessment of particular learning areas or educational levels. For example at ISCED 1: mathematics in 2002 and 2009; Dutch in 2002 and 2007; environmental studies in 2005 and 2010; French in 2008; ICT in 2012.
Canada	SAMPLE – Year 8: mathematics, reading and science on a 3-year cycle with the focus on one of these subjects.
Chile	FULL COHORT – Year 4: language and communication, mathematics, comprehension of natural, social and cultural environments; Year 8 (tested on alternate years): Spanish, communication and mathematics; Year 10: mathematics, language, science, social studies; Year 11: foreign language
Czech Republic	FULL COHORT (under development) – Years 5 and 9: Czech language, foreign language and mathematics.
Denmark	FULL COHORT – Years 2, 4, 6 and 8: Danish; Years 3 and 6: mathematics; Year 7: English; Year 8: geography, biology and physics/chemistry
Estonia	SAMPLE – Years 3 and 6: Estonian, mathematics, science, social studies, technology, modern foreign languages and the arts
Finland	SAMPLE – Most regularly in Year 9, but also in Years 3, 5 and 7: longitudinal assessment plan in mathematics and mother tongue; other subjects tested periodically include science, foreign languages, the second national language, religion, music, art and crafts. Years 3, 6 and 9: cross-curricular learning-to-learn skills. End ISCED 2: communication skills. At ISCED 3 level learning-to-learn skills and communication skills have been assessed once.
France	FULL COHORT – Years 2 and 5: French and mathematics SAMPLE – Years 5 and 9, six subjects assessed alternately on a 6-year cycle
Germany	SAMPLE – Years 3, 8 and 9: German and mathematics; plus modern foreign languages in Years 8 and 9
Greece	None
Hungary	FULL COHORT – Years 6, 8 and 10: national assessment of basic competencies (reading and mathematics). Full cohort since 2007/08. DIAGNOSTIC – Year 4 non-compulsory diagnostic assessment offered in Hungarian and mathematics
Iceland	FULL COHORT – Years 4, 7 and 10: Icelandic, mathematics and science; plus English in Year 10
Ireland	SAMPLE – Years 4 and 8: English reading and mathematics on a 5-year cycle. Prior to 2009, students in different year groups were assessed and not on a specific cycle NO FULL COHORT: But standardised tests (two sets by different test developers) are available in English, Irish and mathematics for Years 4, 6 and 8. The 2011 National Literacy and Numeracy Strategy requires schools to report aggregate student results in these tests to their boards. These will be compiled to report national trends every two years. From 2014, compulsory standardised tests in Year 10 in English reading and mathematics in all schools and in Irish reading in Irish-medium schools. From 2016, compulsory standardised tests in science.
Israel	FULL COHORT – Year 2: language; Years 5 and 8: mathematics, language, science and English SAMPLE – Years 4 and 5: religion
Italy	FULL COHORT – Years 2, 5, 6, 8 and 10: Italian and mathematics
Japan	FULL COHORT – Years 6 and 9: Japanese and mathematics SAMPLE – Periodically there have been national assessments of academic ability, e.g. in 2007 this looked at knowledge and practical use (application) or Japanese and arithmetic/mathematics
Luxembourg	FULL COHORT – Year 5: German and mathematics; Years 8 and 10: German, mathematics and French.
Korea	FULL COHORT – Years 4, 5, 6, 9 and 10: Korean, mathematics, science, social studies and modern foreign languages DIAGNOSTIC – Years 7, 8 and 9: Korean, mathematics, science and social studies

Table 8.3 National assessments in OECD countries, 2012 (*continued*)

	National assessments (excluding international student assessments)
Mexico	SAMPLE – Years 3, 6, 9 and 12: Spanish, mathematics, natural sciences and social studies (each year level tested on a 4-year cycle) FULL COHORT – Years 3, 6, 7, 8 and 9: curriculum knowledge in Spanish, mathematics and one other subject on rotating annual basis (e.g. science, civics and ethics, history, geography). Used by Mexican states to compare student performance.
Netherlands	SAMPLE – Year 5: language and mathematics; Year 8: mathematics/arithmetic. Language, world orientation (social science), English, musical education, physical education, traffic education and visual arts.
New Zealand	SAMPLE – Years 4 and 8: four blocks of curriculum areas assessed alternately on a 4-year cycle (1: science, visual arts, information skills [graphs, tables, maps, charts, diagrams]; 2: language [reading, speaking], aspects of technology; music; 3: mathematics, social studies, information skills [library, research]; 4: language [writing, listening, viewing]; health and physical education).
Norway	FULL COHORT – Years 5 and 8 (since 2007) and Year 9 (2010): basic skills in reading (Norwegian) and mathematics designed to measure cross-cutting competencies. Also skills in English.
Poland	FULL COHORT – Year 6: cross-curricular skills (reading, writing, reasoning, use of information and use of information in practice).
Portugal	None. Until 2012/13 there were full-cohort assessments in Portuguese and mathematics at Year 4.
Slovak Republic	FULL COHORT – Year 9: Slovak language (Hungarian or Ukrainian in ethnically mixed schools) and mathematics. Plans to introduce these in Year 6 also.
Slovenia	FULL COHORT – Year 9: Slovene language (Hungarian or Italian language in ethnically mixed areas), mathematics and a third subject that is determined by the minister for education and sport (taken in four compulsory school subjects taught in Years 8 and 9). Year 6 (schools participate on voluntary basis): Slovene language (Hungarian or Italian language in ethnically mixed areas), mathematics and a foreign language (English or German).
Spain	New national assessment being developed to align with new educational programmes Until 2009 (ISCED 1 – Year 4) and 2010 (ISCED 2 – Year 8) the General Diagnosis Assessment was a full-cohort assessment in Spanish, mathematics, science, social studies, technology, modern foreign languages and the arts
Sweden	FULL COHORT – Years 3, 6 (as of 2012, previously Year 5): Swedish, Swedish as a second language, mathematics and English (Year 6 only). Diagnostic and formative purposes; Year 9 and upper secondary school: Swedish, Swedish as a second language, mathematics and English. Year 9: one science subject (biology, physics, chemistry) as allocated by the National Agency for Education. Summative purposes.
Turkey	SAMPLE – Years 4, 5, 6, 7 and 8: Turkish, mathematics, science, social studies and modern foreign languages.
United Kingdom (England)	FULL COHORT – Years 2 and 6: English and mathematics SAMPLE – Year 6: science
United Kingdom (Northern Ireland)	FULL COHORT – Years 4, 7 and 10: Communication and Using Mathematics (2012/13 onwards) and Using ICT (2013/14 onwards). A mixture of teacher-developed and centrally developed tasks DIAGNOSTIC – centrally developed computer-based adaptive tests are offered to schools, but results are not collected centrally
United Kingdom (Scotland)	None
United States	SAMPLE – Years 4 and 8: reading and mathematics (every two years), science and writing (every four years); also periodic tests in US history, geography, civics, arts and economics; Year 12: reading, mathematics, science and writing.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Many systems choose to administer only sample-based assessments and not full-cohort assessments (the Flemish Community of Belgium, Canada, Estonia, Finland, Germany, the Netherlands, New Zealand, Turkey and the United States). In all of these countries, school participation in the national assessment is voluntary. (Note that sub-national education systems within these systems may administer full-cohort assessments. This is the case for example in some provinces and territories in Canada and the states in the United States). Sample assessments test a representative sample of students and have the major purpose to provide information at the education system level. They aim to measure the overall understanding/mastery of different aspects of the curriculum and not

to assess the performance of individual students. These are, therefore, low-stakes tests. To the greatest extent, they aim to maintain consistent content over time, but would follow – and aim to track success of – changes in curriculum (see examples in Box 8.5). Green and Oates (2009) note that such tests are comparatively low cost and offer other advantages by: providing stable measures to allow the robust measurement of standards over time; allowing a broader coverage of the curriculum; and avoiding distortion of results deriving from "teaching to the test". This latter point is echoed in a report on the accountability system in England within the United Kingdom (House of Commons Children, Schools and Families Committee, 2008). Indeed, one method to increase the curriculum coverage is to use a matrix sampling method (i.e. different content is included in a set of test papers so that not every child sits the same test, thus allowing a broader coverage of topics in the test). The matrix sampling method increases the validity of an assessment as it allows for a more comprehensive evaluation of student performance without increasing student testing time – and also reduces the potential loss of instructional time due to sitting the test (Le and Klein, 2002). While sample-based assessments are favourable if the primary goal is to provide information for system evaluation and related policy making, they cannot be used as a way to identify all schools with performance concerns (Greaney and Kellaghan, 2008).

Such assessments are also used to monitor the performance of sub-systems. Although, the stakes are low for individual students and schools, in Canada and the United States the results from these assessments are used to monitor the validity of average results in state, province or territory assessments – assessments which do carry high stakes for schools. For example, in the United States the National Assessment of Educational Progress (NAEP) aims to provide nationally comparable measures of progress in core skills over time. However, since the introduction of the No Child Left Behind policy requiring individual states to assess students, the NAEP has an additional purpose "to help the U.S. Department of Education verify the results of state-wide assessments" (U.S. Department of Education, 2002). The use of low stakes benchmarking tests to validate tests used in accountability systems becomes increasingly important as stakes are increased in accountability systems (Koretz, 2010). Similarly, in Canada, the Pan-Canadian Assessment Programme is a tool for "Ministers to judge curriculum and other aspects of school systems and to enable provinces/territories to validate/improve their own assessment systems" (CMEC, 2008).

However, many countries choose to administer full-cohort assessments. In some cases this is in combination with a national sample assessment (Australia, Israel, Mexico and England in the United Kingdom). In Australia and Mexico full-cohort testing was introduced in numeracy (Australia)/mathematics (Mexico) and the language of instruction (although in Australia the emphasis is on cross-cutting core skills and in Mexico it is on curriculum content). A full-cohort assessment allows the comparison of how schools perform within the education system, also. Typically, this introduces the additional purpose of holding schools accountable and may include publication of school results, although this is not always the case (see Chapter 6). It also provides comparable data at the level of different sub-national education systems (e.g. municipal or district level, Catholic school sector, etc.) and thus can feed into sub-system monitoring and evaluation. Disadvantages associated with full-cohort assessments are that they can lead to unfair school ranking and cheating and test manipulation by school leaders or teachers (Greaney and Kellaghan, 2008). However, these potential disadvantages are linked to how the test results are used and the extent to which they are perceived as high stakes by educators (Rosenkvist, 2010; Morris, 2011). In

Hungary, the national assessment of basic competencies was extended to a full-cohort assessment in 2007/08. This reflects an increased policy focus on the accountability of individual schools and the attempts to stimulate school self-evaluation activities. Prior, to the introduction of the national assessment of basic competencies in 2001, there was a sample assessment run every two years between 1991 and 2001. The coverage of the national assessments was increased gradually from its introduction in 2001 where it included 20 students from each school to all students in all schools. Similarly in Japan, a sample assessment of Japanese and mathematics was administered in 2007 and this was then introduced as full cohort. Prior to this, there was a history of periodic national sample assessments in Japan.

Box 8.5 Sample surveys in the Netherlands and New Zealand

In the Netherlands, a new monitoring survey – the Annual Survey of Educational Levels (JPON) – was introduced in 2008 to specifically monitor progress on the roll out of the Ministry for Education, Culture and Science's quality agenda "Schools for Tomorrow" and monitors student mastery of Dutch language and mathematics at two points in primary education (Years 4 and 8). Results are reported and analysed for four major regional groupings in the Netherlands. Analysis of performance in urban and rural classifications is also possible (CITO, 2009).

This comes in addition to the existing monitoring sample survey that has been administered periodically in different disciplines since 1987 and monitors skills in Dutch and mathematics on a five-year cycle (Periodical Survey of Education [PPON]). Other curriculum areas that are monitored in the PPON include world studies, history, geography, biology, physics/engineering, English, music and physical education (CITO, 2008). The design of the PPON aims to provide robust measures of changes over time covering large amounts of the curriculum. The design of JPON aims to provide more regular and timely feedback on a narrower area corresponding to the national reform agenda in primary education. Both the PPON and JPON monitoring surveys use Item Response Theory and therefore allow reporting of what students can or cannot typically do against defined performance standards.

In New Zealand, the National Education Monitoring Project (NEMP) established in 1993 assesses students in primary education in two different year groups (Years 4 and 8) and follows a set four-year survey cycle. In this way the NEMP is conducted each year, but assesses a different set of disciplines. For example, in the second year of the survey cycle, music, technology, reading and speaking are assessed, and in the fourth year of the survey cycle, listening and viewing, health and physical education, and writing are assessed. These disciplines, therefore, will only be tested every four years. This allows monitoring of a broad coverage of the national curriculum. According to the NEMP website, the purpose of monitoring samples of students at successive points in time is to identify and report trends in educational performance, to provide good information for policy makers, curriculum specialists and educators for planning purposes and to inform the public on trends in educational achievement.

Sources: CITO (2008, 2009); New Zealand's National Education Monitoring Project website, *http://nemp.otago.ac.nz*.

Capitalising on technology for national assessment administration

In general, national assessments are still predominantly paper based. However, in comparison to national examinations, the administration of national assessments is more likely to capitalise on information technology. This reflects the longer established tradition of national examinations in many education systems. Computer-based uniform technology is used in the administration of national assessments in primary and lower

secondary education in the Czech Republic, Luxembourg (lower secondary only), New Zealand, Norway and for some parts of the national assessment programme in the Flemish Community of Belgium. Computers are also used to administer national assessments to some students with special educational needs in Australia and Slovenia (although for all other students assessments are paper based).

In Norway, the national assessments in English and mathematics have been administered electronically since 2009, but the national assessment in reading (Norwegian) remains paper based. In Denmark computer-based adaptive testing (i.e. the test adapts to the individual student's level as the student takes the test) is used in national assessments in primary and lower secondary education. Similar adaptive testing has been developed in Northern Ireland in the United Kingdom, although these are offered for diagnostic use by schools and the results are not collected for system evaluation. Clearly, systems that have more recently developed national assessments have decided to draw on technology when developing these. National assessments are currently being introduced in the Czech Republic and were relatively recently introduced in Denmark (2009), Luxembourg (2009), Norway (2004) and Australia (2008 for the full-cohort national assessment). In Luxembourg, the University of Luxembourg receives national research funding to develop a computer-based assessment of complex problem solving. Such research can inform the future development of the national standardised assessments (Shewbridge et al., 2012).

Technology can significantly increase the usability and relevance of national assessment results for educators and school managers, as this can dramatically decrease the time required to feed results back to schools. The national assessment system in Denmark delivers test results to teachers, schools and school managers the day after the student has taken the test. Feedback from educators indicates that an increasing majority are reporting using these results in their instructional activities (Danish Ministry of Education and Rambøll, 2012).

Prioritising reliability of national assessments and striving to increase validity

In many systems, national assessments are used for accountability of schools to the public. In this context, the reliability of the assessment is of key importance. This is concerned with the extent to which an assessment is consistent in measuring what it sets out to measure. A highly reliable assessment ensures that the result is accurate and not influenced by the particular assessor or assessment situation. In other words, a highly reliable test if repeated would produce the same results (Harlen, 2007). However, the way in which test results are used within an accountability system also implies challenges to the validity of the tests (see Box 8.8). Validity relates to the appropriateness of the inferences, uses and consequences attached to assessment. A highly valid assessment ensures that all relevant aspects of student performance are covered by the assessment. A need for high reliability and the way in which test results are used both have implications for the design of assessments. Much of the literature implies that there is a trade-off in choosing either a highly reliable test or a test with high validity (see Morris, 2011), but there are ways to better balance both, e.g. using a matrix sample assessment (see above) and providing ways to increase reliability in scoring of tests items that are not closed-format (see below).

As can be seen in Table 8.3, national assessments typically test limited content. With assessments aimed at the curriculum, the major focus is on mathematics and the language of instruction and assessment is often limited to these areas. In other systems, the focus is

on core competencies or cross-cutting skills within the broad categories of literacy and numeracy. Only a few systems test beyond these limited content areas (Korea, Denmark and Spain, notably) and the most typical additional content areas are science and an additional language (foreign language or second language of instruction, as is the case in Luxembourg). In Sweden, national assessments strive to be as well aligned as possible to both academic and democratic goals in the syllabi by incorporating more cultural goals in indirect ways, e.g. through the use of context of testing items and including attitudinal components (Nusche et al., 2011a).

Table 8.4 presents an overview of testing formats used in national assessments, showing the most common format in the first column and the least common format in the final column. This clearly shows that the predominant testing format used in national assessments is multiple choice (i.e. students have to choose one answer from a set of suggested answers), followed by the use of closed-format short answer questions (e.g. yes/no; true/false; selecting a word; providing the result to a calculation). Both multiple-choice and closed-format short answer questions are less costly to develop, administer and score, plus scoring is more reliable and therefore test results are very comparable and can be used to test a range of outcomes (Hamilton and Koretz, 2002; Anderson and Morgan, 2008; and Zucker, 2003, in Morris, 2011). However, in terms of student assessment these formats are associated with testing limited skills and encouraging superficial learning (see Chapter 4; also Morris, 2011). In Mexico, the full-cohort assessments rely entirely on multiple-choice questions.

It is not common practice to include oral questions and answers in national assessments, although this is done in Austria, Iceland, New Zealand and Sweden, as well as in the Flemish Community of Belgium on some occasions. Both New Zealand and Sweden use students' teachers in the scoring of assessment results, but in the other countries, students' oral tasks are centrally marked. Austria and Sweden also include oral presentations in their national assessments. Achieving consistent scoring among different individuals scoring the items or tasks in the tests (or high inter-rater reliability) is key to ensuring the test reliability. This can be aided by the development and provision of scoring guides and rubrics, but also by training the scorers (Shewbridge et al., 2012; Le and Klein, 2002). While scorer training implies higher costs, it can be a good source of professional development (see also Chapter 4).

Multiple-choice tasks and closed-format short answer questions are heavily used due to their higher scoring reliability. That is, there is an obvious right or wrong answer. These can also be automatically scored. However, it is also crucial to ensure reliable administration of the national assessment. With the introduction of national assessments in the Slovak Republic, the State School Inspectorate has been tasked with monitoring the administration of national tests. It visits schools or districts where concerns over test administration have been raised, but also conducts random visits to schools (Hajdúková et al., forthcoming).

Table 8.4 Testing formats used in national assessments (2012)

	ISCED level	Multiple choice	Closed-format short answer questions	Open-ended writing tasks/ calculations	Performing a task	Oral questions and answers	Oral presentation
Austria	1; 2	Yes	Yes	Yes	No	Yes	Yes
Australia	1;2	Yes	Yes	Yes	No	No	No
Belgium (Fr.)	1[1];2;3	Yes	Yes	Yes	Yes	No	No
Belgium (Fl.)[1]	1;2;3	Yes	Yes	No	Yes	No	No
Canada	2	Yes	Yes	Yes	No	No	No
Chile	1;2;3	Yes	No	Yes	No	No	No
Czech Republic	1;2	Yes	Yes	No	No	No	No
Denmark[1,2]	1;2	Yes	Yes	No	No	No	No
Estonia	1	Yes	Yes	Yes	No	No	No
Finland[1]	1;2	Yes	Yes	ISCED 2 only	No	No	ISCED 2 only
France	1;2	Yes	Yes	Yes	No	No	No
Hungary	1[1];2;3	Yes	Yes	No	No	No	No
Iceland	1;2	Yes	No	ISCED 2 only	ISCED 1 only	ISCED 2 only	No
Ireland	1	Yes	Yes	No	No	No	No
Israel	1;2	Yes	Yes	Yes	No	No	No
Italy	1;2;3	Yes	Yes	Yes	No	No	No
Korea	1;2;3	Yes	Yes	No	No	No	No
Luxembourg	1;2	Yes	No	No	ISCED 1 only	ISCED 1 only	No
Mexico (sample)	1;2	Yes	Yes	Yes	No	No	No
Mexico (full)	1;2;3	Yes	No	No	No	No	No
Netherlands[1]	1;2;3	Yes	No	Yes	No	No	No
New Zealand[1]	1;2	Yes	Yes	Yes	Yes	Yes	No
Norway	1;2	Yes	Yes	No	No	No	No
Poland	1	Yes	No	Yes	No	No	No
Portugal	None	a	a	a	a	a	a
Slovak Republic[4]	2	Yes	Yes	No	No	No	No
Slovenia	1;2	Yes	No	Yes	No	No	No
Spain[5]	1;2	Yes	Yes	Yes	No	No	No
Sweden	1;2;3	Yes	Yes	Yes	ISCED 1 only	Yes	Yes
United Kingdom (Northern Ireland)[3]	1;2	m	m	m	m	m	m
Total number of systems using this format:		27	21	17	7	6	3

Notes: The symbol "a" denotes that this is not applicable and the symbol "m" denotes that information is missing. (1) National assessments are not compulsory. (2) Matching items are also used which require students to match pictures/drawings with words. (3) A mix of teacher-developed and centrally developed tasks. (4) National assessments are being developed for ISCED 1 also. (5) Last administered in 2010 and currently under development.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Specific reviews to inform the evaluation of the education system

There are several different mechanisms used by OECD countries to review different aspects of the education system. This may entail the external review of the education system or parts of it. Indeed, this OECD Review offers one example of the latter. Such reviews are conducted to provide objective analysis to inform system improvement. For example, the Northern Territory in Australia commissioned a structural review of its Department of Education and Training in 2009 in order to "more efficiently deliver the (Australian) government's commitments to improved school attendance and levels of literacy and numeracy and meet future challenges in education" (DEEWR, 2010). Such reviews are often stimulated by performance concerns in particular areas. In Denmark, the prime minister commissioned a special review of the public compulsory school (*Folkeskole*) in 2010 following the perceived "mediocre" results in the OECD's PISA studies. Similarly, Queensland in Australia commissioned a performance review in 2009 following results from the Australian national assessments and the IEA's TIMSS study. The purpose of the review was to evaluate the Queensland primary school system and to "identify areas which would potentially lift the educational outcomes for students in Queensland" (DEEWR, 2010). Indeed, in Korea, education system evaluation is conceptualised as the review of specific problematic areas within the education system in order to prepare a plan for improvement and revise the system accordingly. Different concerns may be raised by various stakeholders, including central and local authorities, researchers, teacher and parent associations and the media. A specific committee is assembled to address these concerns. The committee aims to "objectively diagnose the current status of education, clarify the essence of the problem, set standards to measure how the problem is being recognised, research related Korean and foreign precedents, and review and propose improvement plans" (Kim et al., 2010).

Feedback from the external evaluation of individual schools also provides valuable information for education system evaluation. In its simplest form, this may comprise an overall assessment of quality within the education system contained in an annual report from the external school evaluation body, e.g. school inspectorate or school review body (see Table 8.9). However, central authorities may require external school evaluation bodies to conduct reviews in priority areas. This can either be integrated into regular external school evaluation procedures (e.g. the inspectorate or review body examines priority areas in all schools it visits, in addition to areas included in the external school evaluation framework) or in specific "thematic" external reviews. The latter would typically involve a short, focused visit to a sample of schools to examine a specific area.

For example, in the Czech Republic and the Slovak Republic, the school inspectorates have been given the additional task to monitor the extent to which school programmes reflect requirements in national educational programmes. This aims to monitor the implementation of an educational reform and therefore involves an approach to conduct timely reviews in schools. The Swedish Schools Inspectorate conducts thematic reviews as a complement to the regular supervision of individual schools. These comprise around 90% of the Swedish Schools Inspectorate's annual costs. For example, recent thematic quality evaluations have included "Follow-up and evaluation of teaching and learning results", "Bullying, harassment and discrimination in schools", as well as teaching in mathematics, Swedish, physics and modern languages. Similarly, in Ireland, the Inspectorate uses thematic and composite evaluative approaches. Thematic evaluations are specialist evaluation projects with a research focus that examine the quality of selected educational programmes or services in a sample of schools. The approaches involved include first-hand observation of teaching and learning, analysis of documents and work samples, and the use

of structured interviews, focused seminars and questionnaires. Inspectorate composite reports are based on an analysis of subject inspections at secondary level or other inspection data. These reports provide advice and support for teachers, schools and teacher educators, and identify trends and issues of relevance to policy makers.

The specific review of policy implementation can be required as part of external school evaluation (as above), but also can be commissioned to various independent entities, including specific agencies for evaluation and the research community. Although Denmark does not have a system of external school evaluation, there has been the use of external review to monitor the implementation of policies introduced in 2006 to promote and develop an evaluation culture in schools. These have typically been conducted by the Danish Evaluation Institute and the results are used to monitor and adjust policies to ensure more effective implementation (Shewbridge et al., 2011). In a similar vein, Portugal commissioned the University of Lisbon to conduct a review on school implementation of specific national policies, including school use of monitoring and development plans. The results of these and other evaluations show an emerging evaluation culture in Portuguese schools (Portuguese Ministry of Education and Science, forthcoming). In Luxembourg, the Minister may commission the Agency for the Development of Quality in Schools (ADQS) (within the Ministry) to evaluate pilot projects within the education system and recent examples include the evaluation of new types of secondary schools.

International surveys

Countries participate in international surveys with the aim of providing comparative information of key areas in schooling. This typically involves student assessment in core knowledge, skills or competencies (reading, mathematics, science), but can also involve surveys of professionals on teaching and learning (see Box 8.6). Comparative data from international assessment programmes provide a frame of reference that "assists countries in identifying their strengths and weaknesses, provides them with an opportunity for a better understanding of their own system, and offers ideas for further research and policy development" (Tamassia and Adams, 2009). The genesis of international studies of student achievement is the International Association for the Study of Educational Achievement's (IEA) studies which started in the 1960s. Country participation in international achievement studies grew steadily over the years. By the late 1990s all OECD countries had participated in an international study (Table 8.5). The growing importance of international comparisons may reflect both a more outward looking approach to educational policy development from countries and a competitive spirit (Sauvageot, 2008).

Table 8.5 OECD country participation in international achievement studies (2012)

Study first joined	First-time participants
1963-67 First International Mathematics Study (FIMS)	Australia, Belgium, Finland, France, Germany (FRG), Japan, Netherlands, Sweden, United Kingdom (England and Scotland), United States
1968-72 First International Science Study (FISS)	Hungary, Italy, New Zealand
1977-81 Second International Mathematics Study (SIMS)	Canada (British Columbia and Ontario), Luxembourg, United Kingdom (Wales)
1982-86 Second International Science Study (SISS)	Canada (other provinces), Korea, Norway, Poland
1993-97 Third International Mathematics and Science Study (TIMSS 1995)	Austria, Czech Republic, Denmark, Greece, Iceland, Ireland, Mexico, Portugal, Slovak Republic, Spain, Switzerland
1997-2001 Third International Mathematics and Science Study Repeat (TIMSS-R 1999)	Chile, Turkey
2000 Programme for International Student Assessment	All OECD countries, except Chile and Israel (PISA+ in 2002), the Slovak Republic and Turkey (PISA 2003), Estonia and Slovenia (PISA 2006)

Source: International Association for the Study of Educational Achievement and OECD websites.

The OECD launched the Programme for International Student Assessment (PISA) in 2000 and all OECD member countries at that time participated except the Slovak Republic and Turkey (both participated from 2003 on). Although not members of the OECD at the time, Chile and Israel participated in the repeat of the original PISA survey in 2002.

Box 8.6 The major international surveys providing information to evaluate the education system

International Association for the Study of Educational Achievement (IEA)

Participation in different study cycles allows countries "to measure progress in educational achievement in mathematics, science and reading comprehension" and also to monitor "changes in the implementation of educational policy and identification of new issues relevant to reform efforts" (see *www.iea.nl/mission_statement.html*).

- Mathematics and science achievement in Grades 4 and 8: the Trends in Mathematics and Science Study (TIMSS) conducted in 1995, 1999, 2003, 2007 and 2011
- Reading achievement in Grade 4: the Progress in International Reading Literacy Study (PIRLS) conducted in 2001, 2006 and 2011.
- Studies on information technology in education (1997-99, 1999-2002 and 2006) and civic education (1994, 2002 and 2010).

Organisation for Economic Co-operation and Development

Participation in the Programme for International Student Assessment (PISA) provides countries with: a profile of knowledge and skills among 15-year-olds; contextual indicators relating results to student and school characteristics; a knowledge base for policy analysis and research; and trend indicators showing how results change over time, once data become available from subsequent cycles of PISA. Key features of PISA include its policy orientation, innovative approach to literacy and focus on the demonstration of knowledge and skills in a form that is relevant to everyday life.

- Reading, mathematics and science literacy for 15-year-old students: PISA conducted in 2000, 2003, 2006, 2009 and 2012.

Effective teaching and teachers are key to producing high performing students. The Teaching and Learning International Survey (TALIS) offers an opportunity for teachers and school principals to give their input into education analysis and policy development in some key policy areas. Cross-country analysis from TALIS allows countries to identify other countries facing similar challenges and to learn from other policy approaches.

- The learning environment and working conditions of teachers in schools: TALIS conducted in 2008 with a focus on lower secondary school teachers. Next survey in 2013 will allow option to survey also teachers in primary and upper secondary schools.

European Commission

In 2010/11, the first European Survey on Language Competences (ESLC) was conducted in 16 education systems in Europe. Participants gained information on the profiles of student proficiency in the two most widely taught foreign languages in each system (from one of the following EU languages: English, French, German, Italian or Spanish) at either the end of lower secondary (ISCED 2) or the second year of upper secondary (ISCED 3). Contextual questionnaires also provide information on the context of foreign language teaching policies and foreign language learning at student, teacher and school levels.

Sources: IEA website, *www.iea.nl*; OECD PISA website, *www.oecd.org/pisa*; European Commission website, *www.ec.europa.eu/languages/eslc/*.

In Luxembourg, student competencies in French, German and Luxembourgish are key priorities for the education system. As such, Luxembourg participates in a special European project to study language skills assessment (French, German and English) – the European Bank of Anchor Items for Foreign Language Skills (EBAFLS) (Shewbridge et al., 2012). This is conducted with six other countries (France, Germany, Hungary, Spain Sweden and the United Kingdom) and aims to develop assessment items for "reading" and "listening" skills in each language as defined in the European Language Framework (levels A2 and B1). The goal is to make language assessment more transparent and robust.

International assessments offer a rich set of benchmarking data to participating countries. In many systems, these have introduced innovative testing designs and item formats. For example, PISA uses open-constructed response formats that are often not included in national assessments (see Table 8.4). Major technological and methodological advances in international assessments have also raised the prominence of international comparative data in decision making (Tamassia and Adams, 2009). However, a shortcoming of international assessments is their lack of a longitudinal component. Their cross-sectional nature (i.e. one measurement at one point in time) means that it is not possible to measure student progress for a given student cohort (e.g. OECD, 2001; OECD, 2010a; Egelund, 2008; Goldstein and Thomas, 2008). This also calls for caution on using such correlational data to infer causal relationships. International assessment results cannot provide definitive evidence on performance variations, but can raise important questions to be investigated by further research (Goldstein and Thomas, 2008; McGaw, 2008). These results can inform policy makers of the need for other national studies. Caution should be taken in interpreting results due to the "often narrow focus of comparative data and the consequential risk of misinterpretation" (Tamassia and Adams, 2009). The use of an international framework limits the ability of international assessments to give a comprehensive overview of any one national system (Bialecki et al., 2002; Ofqual, 2008). Finally, the international assessment cycle may not come at optimum times to monitor extensive system reform or innovation (Green and Oates, 2009; Ofqual, 2008).

Key stakeholder surveys

Countries may administer surveys to key stakeholders to collect more qualitative feedback on the education system, including on the teaching and learning environment, overall satisfaction and maybe on specific innovations to the education system. Table 8.6 provides an overview of the use of stakeholder surveys by OECD countries. It is most common among OECD countries to collect qualitative feedback from students. Many countries include short questionnaires for students as part of their national assessments. This provides useful feedback in its own right and can also be used to analyse how certain factors relate to performance. However, specific feedback surveys for students may be used independently of national assessments. For example, Norway administers an annual student survey and reports major results and analysis as part of its annual summative report on the education system (see Box 8.11). In Australia, a National School Opinion Survey collecting responses from students and parents will be implemented from 2013. It is anticipated that a similar teacher survey will also be implemented in 2015. Results of the surveys are to be published in the school's Annual Report.

Teacher surveys are administered in Australia, Austria, Chile, Estonia, Iceland, Israel, New Zealand, Norway, Portugal, Sweden and Northern Ireland in the United Kingdom, as well as in the Flemish Community of Belgium, Canada and Hungary as part of the

national sample assessment. With the exception of Canada, Estonia, Israel and Northern Ireland in the United Kingdom, all these countries also administer parent surveys. Parent surveys are also used in Ireland, Italy, the Netherlands and Poland. In France, the Federation of Parents of Public School Students (PEEP) regularly conducts parent surveys, although there are no officially developed parent surveys. Several countries also participate in the OECD's Teaching and Learning International Survey (TALIS) (see above).

In Luxembourg, although there are no regular stakeholder surveys, part of the mandate for the Agency for the Development of Quality in Schools (ADQS) is to collect and synthesise qualitative feedback from schools, as commissioned by the Minister (Shewbridge et al., 2012). This sends a strong signal on the importance of collecting feedback from key stakeholders. Recent evaluations of national school innovation projects and the implementation of other national policies have used teacher and parental surveys to seek feedback. The results of these opinion surveys are synthesised and analysed centrally and they feed into the considerations for further development and refinement of the specific policies.

In Denmark, there are no regular stakeholder surveys, although a central framework for collecting feedback from parents has been developed and local authorities are free to use this. Copenhagen runs an annual student survey. Analysis of results from the Copenhagen student survey shows strong association between student performance and many qualitative aspects of school life. This indicates that the collection of such data nationally could be of significant policy and research interest (Shewbridge et al., 2011).

Table 8.6 The use of stakeholder surveys in the OECD (2012)

Stakeholders	Countries administering surveys
Students, parents and teachers	Australia, Austria, Flemish Community of Belgium[1], Chile, Hungary[1], Iceland, New Zealand[1], Norway, Portugal, Sweden
Students and parents	Ireland, Italy, Netherlands, Poland
Students and teachers	Canada[1], Estonia, Israel, Northern Ireland (United Kingdom)
Students	French Community of Belgium, Finland[1], Spain
No stakeholder surveys	Czech Republic, Denmark[2], France, Korea, Luxembourg, Mexico, Slovenia, Slovak Republic

Notes: (1) In the Flemish Community of Belgium, Canada, Finland, Hungary and New Zealand this is part of the national sample student assessments. In Hungary there may also be ad hoc surveys. (2) In Denmark, there is a central framework to collect feedback from parents which local authorities can choose to use.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

The use of a survey cycle can broaden the range of stakeholders consulted and the coverage of information collected. New Zealand uses national surveys on a 3-year cycle that are completed by a random sample of primary and secondary schools (principals, teachers, trustees and parents) on school finance, strategic management, professional development and collaboration with communities, plus other current priority topics, e.g. the introduction of national standards in 2010 (Nusche et al., 2012). In Norway, the Directorate for Education and Training has established a multi-year framework for administering sample-based user surveys to ensure a cyclical coverage of key topics, while limiting demands on users to complete surveys. The regular collection will allow monitoring and reporting on seven key areas, but will ensure that school principals and local authorities only complete a survey once every 18 months (Norwegian Directorate for Education and Training, 2011).

Graduation and completion data at the end of key stages of schooling

Although, the primary purpose of national examinations is to give a summative assessment of an individual student's achievement, the results from such examinations (and indeed final marks awarded by teachers) are often aggregated to show overall achievement at the national and sub-national levels. These are considered to be key indicators of student outcomes at the end of schooling. The primary purpose of national examinations means that the tests are designed to satisfy criteria for high-stakes assessment. In this context, test content would be less predictable and would not be designed to link results from year to year. These are, therefore, not ideal measures for tracking changes over time. However, a high reliability in scoring and administration conditions across schools would mean that these should provide good measures for comparing student performance at a given point in time (see Chapter 4). As such, the information can feed into system evaluation – although aggregate outcomes in different regions and schools are largely influenced by contextual factors (see Box 8.8).

Longitudinal surveys

Table 8.7 presents an overview of different types of longitudinal information available to OECD countries for education system evaluation. Many systems have longitudinal information on the progress of student cohorts through the education system – this is a relatively recent possibility in some systems corresponding with the introduction of national assessments that are designed to follow student cohorts.

In the Flemish Community of Belgium, longitudinal research is conducted by the Centre of Education and School Careers (*Steunpunt voor Studie-en Schoolloopbanen*, SSL). Research covers both primary and secondary schooling (*www.steunpuntloopbanen.be*). Longitudinal studies of student cohorts are well established in the Netherlands (see Box 8.7). In New Zealand, the "Competent Children, Competent Learners" longitudinal research programme specifically analyses the development of different competencies (knowledge, skills and attitudes) across a student cohort and identifies factors associated with this with the aim to identify promising directions for improving children's competency levels (*www.nzcer.org.nz/research/competent-children-competent-learners*). "Growing Up in Ireland" is a longitudinal research programme that follows cohorts of children and examines factors impacting social, health and emotional outcomes for children (*www.growingup.ie*).

Since 1996, New Zealand has introduced a unique student identifier (the National Student Number, NSN). This can be used for longitudinal research studies. However, student privacy must be respected. This unique identifier facilitates the management and sharing of information about students across the education sector in a way that protects their privacy (Nusche et al., 2012). At the level of the Ministry of Education, almost all data collection from schools is set up in a way as to enable longitudinal analysis, using the NSN as a link. The existence of a widely applied unique identifier covering both schooling and the tertiary sector is a key strength of system monitoring in New Zealand. The NSN can be used by authorised users for the following five purposes: monitoring and ensuring a student's enrolment and attendance; ensuring education providers and students receive appropriate resourcing; statistical purposes; research purposes; and ensuring that students' educational records are accurately maintained. Among other things, the NSN is applied for reporting purposes by education agencies, analysis of student assessment data over time, moving data between software applications, and issuing documentation students need to present to other schools or education providers.

Box 8.7 Longitudinal research programmes in the Netherlands

In the 1980s there was a debate about ways to evaluate ongoing policy programmes in education, e.g. stimulating equity in primary schools and restructuring secondary education. The Foundation for Educational Research (SVO) advised that experimental designs were not feasible and proposed the use of cohort studies, i.e. to analyse the educational attainment of age cohorts of students throughout their school career.

The cohort study is jointly financed by the Ministry of Education and the Foundation for Scientific Research (NWO). The COOL study is presented as a data source for various user groups, such as the Ministry of Education, The Social and Cultural Planning Bureau and the Educational Council (*Onderwijsraad)*. Further, research groups can use the data to conduct more in depth studies. The COOL study also comprises specific services to feed back information to schools.

Initially, two different cohort studies in primary education (PRIMA) and secondary education (VOCL) were launched. Essentially both comprised: achievement testing of students in language and mathematics/arithmetic at various levels during the school career; recording data on students' progress throughout the school programme (class repetition, drop out, transfer to another school type, and examination results); and the collection of school background data by means of questionnaires to parents, teachers and school leaders. Schools providing special education were not included.

From 2007, the primary and secondary cohort studies were integrated into the Cohort Survey School Careers (*Cohort Onderzoek Onderwijsloopbanen*, COOL). The COOL study comprises data collection in primary (Years 2, 5 and 8), general secondary (third year) and vocational schools (second year). COOL surveys have been planned until 2015. Student achievement is tested in language and mathematics/arithmetic. The achievement tests make use of the Cito pupil monitoring system LOVS. In addition short questionnaires are administered to students. A pre-COOL study in early childhood education will also be launched.

See purpose and intent at *www.cool5-18.nl/doelenopzetbo/#4*.

See information for schools at *www.cool5-18.nl/scholenbo/*.

Source: Scheerens et al. (2012).

In France, since 1973 the Education Ministry has been responsible for longitudinal panel surveys of students starting in either primary or lower secondary education (*panels d'élèves*). They aim to follow the pathways and educational development of a cohort of students until their exit from education. Information on the progression of individual students is also collected via a personal competency booklet for each student. Further, there are studies on student entrance into the labour market upon completion of education.

Australia and Canada have linked existing longitudinal studies to follow up students who participate in PISA to shed light on student pathways. The fact that there were existing national longitudinal studies meant that neither country was daunted by the technical and administrative complexity involved (McGaw, 2008).

The Longitudinal Surveys of Australian Youth (LSAY) began in 1995 and collects information from 10 000 young adults once a year over a period of ten years about their education and training, work and social development (more information on *www.lsay.edu.au/aboutlsay/about.html*). For example, information is collected on student achievement, student aspirations, school retention, social background, attitudes to school, work experiences and what students are doing when they leave school (e.g. vocational and higher education, employment, job seeking activity). Surveys in 2003, 2006 and 2009 are integrated with PISA.

As part of the "Youth in Transition Survey", the 30 000 Canadian students that participated in the PISA 2000 survey have been interviewed every two years (final interviews were in 2010 when the young adults were 25 years old) to track their education/work transitions. For example, questions in 2006 related to engagement in and financing of post-secondary education, education and work aspirations, employment status, income, personal characteristics and background. Based on this follow-up, results demonstrate strong association between better performance in PISA and student completion of secondary school and participation in some form of post-secondary education (OECD, 2010c).

Table 8.7 Availability of longitudinal information for education system evaluation (2012)

Availability of longitudinal information	Systems in which this is available
The progression of individual students: The progression of student cohorts; Future studies/professional pathways of graduates	Australia, Canada (this varies among provinces/territories), France
The progression of individual students; The progression of student cohorts	Austria, Slovenia, Finland, Hungary
The progression of student cohorts	Chile (since 2007), Denmark, Italy (since 2008/09), Portugal (since 2007/08)
The progression of individual students	Luxembourg, Mexico, Norway (ISCED 3)
Longitudinal research programmes	Belgium (Fl.), Estonia, Ireland, Israel, Netherlands, New Zealand, Norway, Poland
No longitudinal information	Belgium (Fr.), Czech Republic, Iceland, Korea, Slovak Republic, Spain, Sweden, United Kingdom (Northern Ireland)

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Similarly, in Denmark there was a follow-up of the participants in PISA 2000 using individual registration numbers and personal interviews conducted in 2004 (see Jensen and Andersen, 2006; Egelund, 2008). In Luxembourg, a special follow-up national study (LESELUX) was conducted after the IEA's PIRLS 2006 assessment to more deeply investigate and confirm key messages revealed by the main study (Shewbridge et al., 2012).

Meeting information needs at the system level

A key challenge within the evaluation and assessment framework is to develop indicators and measures of system performance that permit a good understanding of how well the education system is achieving its objectives. While national education goals may be comprehensive and broad, monitoring systems may be rather limited in the information they can offer. This runs the risk of policy being driven primarily in areas where there are measures available. At the same time, heightened demands for information may necessitate significant resources in both the different bodies responsible for collecting and compiling this and schools and other stakeholders that may need to provide this. With regard to quality goals of the education system, some systems may lack quantitative measures in key areas of schooling or rely on a limited set in discrete areas. Other systems may lack qualitative measures in key areas of schooling, including for example, information on the quality of teaching and learning and feedback on the satisfaction of students, parents and educators. In this context, it is challenging to juggle the different tensions between feasibility, reliability, coverage and validity.

One key concern in system evaluation is to monitor equity goals across the system. This means that systems need to ensure the collection of basic demographic data for students and schools. Typically, data of interest for system analysis include: student socio-economic background (often measured by their parents' education level and occupation); student first language and whether this is different from the language of instruction; student place of birth; and information on any special educational needs. Such information can be available via Labour Force Surveys, as well as regular population census and may be collected via the administration of questionnaires to students during national assessments. In the Czech Republic, France, Iceland, Ireland and Sweden, information on socio-economic composition of schools is available, but this is not collected at the individual student level. In Poland, student background information is not systematically collected, but only in specific educational research. In Portugal information on students' socio-economic background at the student and school levels is not collected as part of the regular submission of data from schools in administrative collection, nor via the administration of the national tests. However, for public schools and government-dependent private schools, information is collected on whether or not students receive free school meals and assistance for special educational needs, and also on the profession of the students' parents.

Monitoring key student learning outcomes

In all countries, there is some form of system in place to monitor student learning objectives. However, monitoring systems only provide performance information against discrete parts of student learning objectives, although the breadth of areas covered varies significantly among countries. As student learning objectives evolve this brings new demands for monitoring systems. Standardised national assessments heavily draw on multiple-choice assessment formats and this poses a challenge to monitoring how well students perform in more cognitively demanding areas. In particular, heightened importance in student learning goals for cross-cutting competencies and student ability to demonstrate and apply their learning in real-life settings, may not yet be adequately reflected in national monitoring systems. Conversely, national assessments may be designed to assess student competencies that are not yet adequately reflected in student learning objectives and/or implemented at the school level.

Securing comparability over time and across schools

Countries often face challenges in monitoring trends in student performance over time. This might be because national assessments or examinations are not designed in such a way results are comparable over time. This limits the ability to analyse the progress of student performance over time. Another challenge often concerns the comparison of student outcomes across schools. This might arise because there are inadequate mechanisms in place to ensure the comparability of marking and scoring of results. Further, if results of student assessments are made public at the school level with no account for the socio-economic context of each school (or the characteristics of schools' student population), this can considerably distort considerations about the effectiveness of each school as average results do not reflect the value added by schools to student results (see also Chapter 6). Box 8.8 presents an overview of the methodological concerns in designing appropriate measures.

Box 8.8 Tensions between monitoring changes over time and holding schools accountable

Experts in testing and statistical methodology caution that there are tensions in securing measures that are comparable over time and that provide meaningful comparisons across schools – especially when these are perceived to be high stakes by schools. This, therefore, has implications for system monitoring if the results from full-cohort national assessments are used in school accountability systems. Introducing some measures of accountability may be a way for countries to enforce and support a culture of using student test results for improving instruction (Rosenkvist, 2010). However, introducing a system of accountability requires a clear understanding of what student tests can (and cannot) be used for.

Measures that are best suited to school comparisons are unstable measures of changes over time

A fair and legitimate school accountability system should only hold teachers and schools accountable for factors that they can influence (Rosenkvist, 2010). However, in a review of literature on different types of school performance measures used in accountability systems, Masters (2012) shows that measures that have been developed to provide fairer or more meaningful comparisons of school performance do not lend themselves to monitoring changes over time. Specifically, "gain scores" or "growth" measures (i.e. measuring the same set of students on the same measurement scale at two points in time, e.g. cohort's reading scores between Year 3 and Year 5) and contextual value added measures (CVA) (i.e. an adjustment for contextual influences in a school by investigating the association of a range of contextual factors and student test scores at the education system level). Research has consistently found that gain scores are unstable over time and it is therefore difficult to draw conclusions about changes in gain scores with any degree of confidence. Changes in CVA are very difficult to interpret as measures of school improvement as these have no easily understood substantive meaning. Conversely, "status" measures (e.g. average performance on a test or percentage of students at a particular level of proficiency) are strongly influenced by contextual factors and students' prior learning, so are less meaningful when comparing schools with different intake. However, they can provide useful information on changes over time within the same school, if there is evidence of improvement over a sustained period (given possible fluctuations from year to year with particular student cohorts).

High-stakes use of test results implies changes in test design and validation

Koretz (2010) argues that the prominence of tests in accountability systems necessitates fundamental changes in test design and validation. Even though the shift to high-stakes testing in the United States has led to a number of innovations in testing, these have not been enough and do not address the core concerns. The concern of reallocation within schools to the subjects tested at the expense of other subjects is, Koretz argues, a policy and not a measurement issue. However, the predictability of test content may lead to a reallocation of instructional time within subjects, which is a measurement issue as it can lead to test score inflation. Further, predictability of non-substantive parts of a test item, e.g. its format and task demands, may lead to coaching in form of guidelines emphasising mechanical aspects of test taking. Under high-stakes conditions, predictable tests can lose their value as measures of the broader domain as educators and students have strong incentives to behave in ways that undermine this. As such, Koretz advocates for less predictable tests, but this entails higher test development costs and technical concerns for linking tests over time. This heightens the importance of having an additional low stakes benchmarking test that can be used to validate tests used in accountability systems.

Measures should be fit for purpose and the results and their limitations communicated clearly

Perie and Park (2007) conducted a review of accountability systems and related literature in the United States and caution that a well-designed and effective accountability system pays attention to seven core components: 1) The reason for the accountability system and its intended goals (expected outcomes); 2) Performance indicators used are valid and interpreted correctly and should be as many as possible and not just one measure; 3) Design the system to match the intended goals (e.g. whether to use status, improvement or growth measures); 4) Consequences of the results regarding possible sanctions and rewards and monitoring these for effectiveness; 5) Communication about the accountability system and its results and their limitations to schools, school providers and the general public; 6) Support from the State to schools for improvement and evaluating whether the accountability system supports high-quality instruction; 7) Regular evaluation, monitoring and improvement of the system. For an overview of literature on the importance of maximising test validity by ensuring a use of test results that is fit for purpose see Rosenkvist (2010) and Morris (2011).

Sources: Koretz (2010); Masters (2012); Perie and Park (2007); Rosenkvist (2010).

Capacity

Among OECD countries, there are several different entities with a role in education system evaluation and these can typically be classified into two major groups. Those with capacity to:

- Compile key information for national monitoring: e.g. National Bureau of Statistics or statistical units within the Ministry of Education (internal or outsourced) and students' teachers and schools.
- Produce evidence and use education system evaluation results: e.g. education authorities at the national and sub-national levels; bodies conducting national and international assessments; quality assurance agencies; inspectorates or school review agencies; national audit offices; education councils bringing together a range of stakeholders; research institutes.

Compiling information for national monitoring

Schools are required to submit information on their schools that feeds into both school evaluation and education system evaluation (see Chapter 6 for details on compliancy reporting requirements at the school level). For example, some systems require schools to provide information annually on student demographics and socio-economic characteristics and this forms an important part of the monitoring of equity goals within the education system.

In some systems, students' teachers are engaged in the scoring of their work in national assessments. Teachers bare sole responsibility for this in the French Community of Belgium, Finland, Slovenia (ISCED 1 only) and Sweden. Central scoring guidelines are provided. In the French Community of Belgium and Sweden, this is organised at the school level. The Swedish Schools Inspectorate may also monitor the scoring by teachers in different schools. Results of such evaluations revealed significant differences in scoring practices among schools (Nusche et al., 2011a). In France (ISCED 1 only), the Netherlands and Norway, both students' teachers and central authorities score student work in the national assessments.

Producing evidence and using education system evaluation results

Over recent years, many countries have established specific bodies to oversee quality assurance of the education system. This aims to signal the importance of evaluation within the system also. However, countries draw on different entities for evaluating the education system and sub-national systems (Table 8.8).

Table 8.8 Capacity for evaluating the education system and sub-national systems (2012)

	Conducting evaluations on identified priority areas within the education system	Evaluating sub-national education systems
Australia	Central education authority or government; state education authorities or governments; audit office; research institutes; other (Productivity Commission, external validators)	State education authorities or governments (mainly for government schools only); some systemic non-government school systems conduct system evaluation
Austria	Central Agency (Federal Institute for Education Research, Innovation and Development of the Austrian School System, BIFIE); research institutes); national statistical office; audit office	State education authorities; research institutes (Universities, University College of Teacher Education, other scientific institutions)
Belgium (Fl.)	Central education authority; research institutes; other (umbrella organisations for schools)	a
Belgium (Fr.)	Central education authority; Inspectorate	a
Canada	Provincial/territorial education authorities (ministries of education); Council of Ministers of Education, Canada	Provincial/territorial education authorities (ministries of education)
Chile	Central agency (The Quality of Education Agency)	Central agency (The Quality of Education Agency)
Czech Republic	Central education authority (Ministry of Education, Youth and Sports); Czech School Inspectorate; regional education authorities	Czech School Inspectorate; regional education authorities (14 regional education sub-systems)
Denmark	Central education authority; central agency (Quality and Supervision Agency); research institutes; local education authorities	Central education authority; central agency (Quality and Supervision Agency); research institutes; local education authorities
Estonia	Central education authority (Ministry of Education and Research)	Central agency (Foundation Innove); regional education authorities (county governments); local education authorities
Finland	Central education authority (Finnish National Board of Education); central agency (Education Evaluation Council in Jyväskylä; research institute (Finnish Institute for Educational Research at the University of Jyväskylä)	Local education authorities
France	Central education authority; General Inspectorates (IGEN; IGAENR); research institutes; National Council for Education (advisory body[HCE]); National Audit Offices (*Cour des Comptes*; IGF); Parliament	Central education authority; General Inspectorates (IGEN; IGAENR); regional education authorities (*académies*) with the support of local and regional inspection bodies (IEN; IA-IPR); regional statistical services (SSR)
Hungary	Central education authority; Institute for Educational Research and Development	a
Iceland	Ad hoc independent expert groups; research institutes	a
Ireland	Inspectorate of the Department of Education and Skills; research institutes	a
Israel	Central education authority	Central education authority
Italy	Central evaluation agency (National Institute for the Evaluation of the Education System, INVALSI), National Institute for Educational Research, Experimentation and Development (INDIRE); Inspectorate	a
Korea	a	Central government; research institutes
Luxembourg	Central education authority	a
Mexico	Central education authority; central agency (National Institute for Educational Assessment and Evaluation, INEE)	State education authorities
Netherlands	Central education authority; Dutch Inspectorate of Education; Audit office; research institutes	a
New Zealand	Central education authority or government; central agency (Education Review Office)	Central education authority or government, central agency (Education Review Office)
Norway	Central education authority (Norwegian Directorate for Education and Training)	Local education authorities
Poland	Regional education authorities (educational superintendents)	Local education authorities

Table 8.8 Capacity for evaluating the education system and sub-national systems, 2012 (*continued*)

	Conducting evaluations on identified priority areas within the education system	Evaluating sub-national education systems
Portugal	Central education authority (General Inspectorate of Education and Science; General Directorate for Education and Science Statistics)	These are ad hoc and may be conducted by different central bodies (e.g. National Agency for Qualification and Vocational Education and Training)
Slovak Republic	Central education authority (Ministry of Education, Science and Sport); State School Inspectorate; National Institute for Certified Educational Measurements (NÚCEM); regional education authorities	Regional education authorities (8 regional education sub-systems)
Slovenia	Inspectorate of the Republic of Slovenia for Education and Sport; central agency (National Examinations Centre), research institute	a
Spain	Central education authority; state education authorities	State education authorities or governments
Sweden	Central education authority; central agency (National Agency for Education); Swedish Schools Inspectorate	a
United Kingdom (Northern Ireland)	Central education authority (Department of Education)	Central education authority (Department of Education)

Note: The symbol "a" denotes that this is not applicable.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

In Portugal, while there is no specific body responsible for the overall evaluation of the education system, the General Inspectorate of Education and Science audits and evaluates both the quality of education and teaching within the system and the Ministry of Education and Science. Within the Ministry, a specific department is responsible for developing national student assessments and all departments are responsible for monitoring the areas under their responsibility (Portuguese Ministry of Education and Science, forthcoming). In Luxembourg, a department within the Ministry of Education and Training co-ordinates research on pedagogical and technological innovation. The department includes a specific Agency for the Development of Quality in Schools (ADQS) that was created in 2009 to both provide information to monitor schools and the education system and to support and develop school capacity to conduct self-evaluation. In the Slovak Republic, the Ministry of Education, Science and Sport holds overall responsibility for ensuring the quality of the education system, but delegates authority to two institutions to evaluate the overall performance of the education system: the Slovak State Schools Inspectorate and the National Institute for Certified Educational Measurements (NÚCEM). The latter was established in 2008 and marks a move to strengthen national capacity in monitoring student performance. In Austria, one of the major barriers to implementing a monitoring system in the past was the absence of institutions and infrastructure to conduct cross-system assessments (Specht and Sobanski, 2012). In 2008 a central agency was established to address many of these concerns (see Box 8.9).

In general, similar trends to build capacity in evaluation and quality improvement at the national level can be observed in the OECD Review countries (Box 8.9). However, countries also make use of research institutions and other evaluation bodies in conducting specific reviews (Box 8.10).

In Sweden, the National Agency for Education (NAE) runs a policy to employ a mix of educators, statisticians, political scientists, sociologists, economists, lawyers, etc (Nusche et al., 2011a). There are three distinct units within the NAE: education statistics; evaluation of results (national/international); and analysis and reviews. While the unit for educational statistics processes and conducts quality control on national test data, the NAE commissions Statistics Sweden to collect these data adhering to strict technical requirements. The NAE works with researchers and teacher trainers to develop the syllabi, which serve as the standards in the Swedish education system. Once the syllabi are completed, the NAE starts work on designing related national tests. The NAE manages the national test system and contracts various universities to develop the tests in consultation with representative teacher groups. The NAE also collaborates with universities on the development of other materials such as school quality management tools.

Box 8.9 National agencies with specific responsibility for education system evaluation

Italy

In July 1999, the legal decree 258 established the Italian Institute for the Evaluation of the Education System (INVALSI). This built on existing national capacity in research and evaluation (the European Centre for Education [CEDE]), but significantly increased the importance of the evaluation of the productivity and effectiveness of the education system as a whole. Not exhaustively, INVALSI: manages the National Evaluation System (VNS) and performs periodic audits on the overall quality of schools; develops national assessments and manages Italy's participation in international assessments; provides technical support for evaluation activities at the regional and school levels; provides evaluation-related training for teachers and school leaders; conducts research; and develops support tools for implementing school leader appraisal.

Austria

On 1 January 2008, the National Council established the Austrian Institute for Education Research, Innovation and Development in Schooling (BIFIE), an independent legal entity with clear roles and responsibilities. This represents a significant increase in the volume and quality of education research activities in Austria. BIFIE's mission statement is as follows: "A country's future prospects are inextricably linked to the quality of the education system. To further improve the Austrian education system, it is necessary to take stock of the current situation, to implement effective reforms and then to evaluate these. The basis for this is the development of evidence-based education policy and systematic school development."

Denmark

Since 2006, the creation of a new national structure has aimed to signal and establish the importance of evaluation and assessment in compulsory education, including: an advisory body with representatives from all the major stakeholder groups to inform the Minister of Education on the quality of the compulsory public schooling (*Folkeskole*) (the School Council for Evaluation and Quality Development of Primary and Lower Secondary Education) and an agency to monitor and develop quality in compulsory education (the School Agency). On 1 March 2011, the School Agency was restructured and renamed the Quality and Supervision Agency, but kept many of its previous responsibilities. It is responsible for the financial, institutional and pedagogical supervision of both compulsory and upper secondary education, plus quality development in these sectors. Major tasks include developing and running the national assessments and final examinations in compulsory education, managing the implementation of international assessments, plus the development of evaluation support materials for schools and the dissemination of these via an Evaluation portal. This complements the existing evaluation capacity within the Danish system, i.e. the Danish Evaluation Institute (EVA) established in 1999 to help bring about a shift from a focus on inputs to outputs (OECD, 2004). EVA conducts evaluations in all levels of education in Denmark. Since 2006, its evaluations in compulsory education are commissioned by the School Council.

Ensuring objectivity and credibility in evaluation activities

Education system evaluation results are heavily used by policy makers and also at the highest political levels. To varying degrees among countries and over different political cycles, there may be different tensions put on national evaluation bodies, including limited resources available for their activities, restructuring and in some cases, closure. A commitment to evidence based policy making is commendable and it is essential to pay attention to the objectivity of different evaluations and evaluation bodies. Politically commissioned ad hoc evaluations may run the risk of lacking credibility in the eyes of some stakeholder groups. A national body responsible for education system evaluation can provide technical autonomy from the education authorities with the necessary distance from political decision making to conduct rigorous and reliable analyses of data. A national body can confront the education authorities where necessary and be impartial in its conclusions about the education system. This can provide a fresh and constructive external point of view informing the national debate.

In Mexico, the President proposed in December 2012 a constitutional reform to Congress to create the National System of Educational Evaluation. This will be coordinated by the National Institute for Educational Assessment and Evaluation (*Instituto Nacional para la Evaluación de la Educación*, INEE) which will be an independent public body with its own legal personality and assets. It will be responsible for the evaluation of the quality, performance and results in pre-school, primary, lower and upper secondary education. The INEE was initially created in 2002 as a dedicated body responsible for education system evaluation. Its functions include: development of a national indicator system and learning outcomes assessments; design of evaluation instruments adapted to each level of the education system; collaboration with the Ministry of Education and state governments in assessment and evaluation; development of a school evaluation model; stimulation and strengthening of an evaluation culture; dissemination of results; capacity building at different levels; evaluation of selected projects and development of research regarding evaluation. The 2012 revision to the 2002 Presidential Decree reinforces INEE's autonomy, strengthens its technical expertise, and provides further independence from the Ministry of Education.

In Portugal, the National Education Council (CNE) reviews the implementation of national education policy and provides independent advice to the government on national education issues. The CNE has the right to initiate reviews, investigate specific themes, report and publish findings on the provision of education in Portugal. The reports published by the National Education Council (CNE) add considerably to the national information base which informs policy development. An advantage is that the CNE is in a position to provide a unique input for system monitoring and policy development: views and perspectives of a wide range of education stakeholders. Its recommendations are based on a consensus among its diverse membership, therefore informing policy implementation about the areas of agreement and contention. The thematic reports that it commissions also allow the education research community to contribute to the development of education policy. In Denmark there is a similar advisory body formed of major stakeholders that can commission reviews and report on findings (see Box 8.16).

Box 8.10 Commissioning specific education reviews in Korea and the Netherlands

In Korea, education system evaluation is often conducted by the central government or research institutes. However, specific committees are assembled to conduct evaluation in priority areas that are identified by the public as serious social issues. The committee may comprise educators, experts, journalists, lawyers, parent and teacher associations who will undertake decision-making roles, but also other experts and professional staff who support the evaluation with analysis of evidence and international practices. Each committee prepares an improvement plan and this is submitted to the central government and if necessary may lead to legal revisions or a presidential decree. This approach using "presidential committees" has been used since the 1960s in Korea. However, the improvement and reform plans developed by such committees "were often processed with difficulty, as they met with conflicting interests among various stakeholder groups" (Kim et al., 2010).

In the Netherlands, specific programme evaluations have been conducted by evaluation committees resided directly under the Ministry of Education. However, the Dutch Inspectorate of Education has also been involved in co-ordinating and partially undertaking two major policy evaluations in primary and lower secondary education.

Sources: Kim et al. (2010); Scheerens et al. (2012).

Ensuring adequate evaluation capacity at the sub-national level

In many countries, sub-national education systems play a major role in evaluation activities. For example, there may be national requirements in place to stimulate quality monitoring by local authorities. This aims to stimulate external school evaluation and school self-evaluation processes (see Chapter 6). It is a challenge to ensure that there is enough capacity at sub-national levels to sustain effective quality assurance systems, including robust monitoring of schools. Crucially, local capacity to follow up on evaluation results and to work with schools for improvement often varies significantly within countries and poses significant challenges to smaller local authorities.

In Norway, several local authorities do not have a monitoring structure in place and "are less capable of following up the results of schools" (Norwegian Department of Education and Training, 2011). The Directorate for Education and Training provides capacity building and support offers for local authorities facing challenges related to the quality of their local systems (Nusche et al., 2011). Notably, a "Guidance Corps" of exemplary school principals has been established to intervene in local authorities that have been targeted as needing help with capacity development. Forty local authorities were invited to join a voluntary support programme and 31 decided to participate indicating that this is a welcome initiative. There have also been capacity building initiatives by led by local authorities or the umbrella organisation for local authorities (KS). Since 2005, KS has launched different local authority "efficiency" networks and offers a set of quality monitoring tools (Norwegian Directorate for Education and Training, 2011). In general, networks consist of between four to eight local authorities and run for four to five meetings and then disband. There are many different regional networks, plus one network with the 10 largest local authorities which focus very much on benchmarking style exercises.

In Denmark, there have been reporting requirements since 2006 for local authorities to produce annual quality reports showing an overview of the performance of schools under their authority (see also Chapter 6). The Danish School Agency provides

information exchange among local authorities on their different approaches to using the results of this quality reporting exercise (see Shewbridge et al., 2011). A national electronic portal provides a central reference point to record different municipal approaches to quality assurance and development. Plus, the Danish School Agency has organised conferences to stimulate municipal exchange and partnerships. At the local government level, an umbrella organisation for local authorities (Local Government Denmark, KL) ran a two-year partnership involving 37 local authorities which focused on local quality assurance as one of three priority areas for development and included use of a suite of key indicators – measured via questionnaires administered to school principals, teachers, parents and students at both the start and end of the partnership – to shed light on the impact of the partnership. Results revealed both a greater focus on results and better use of the mandatory local authority quality reports (KL, 2009).

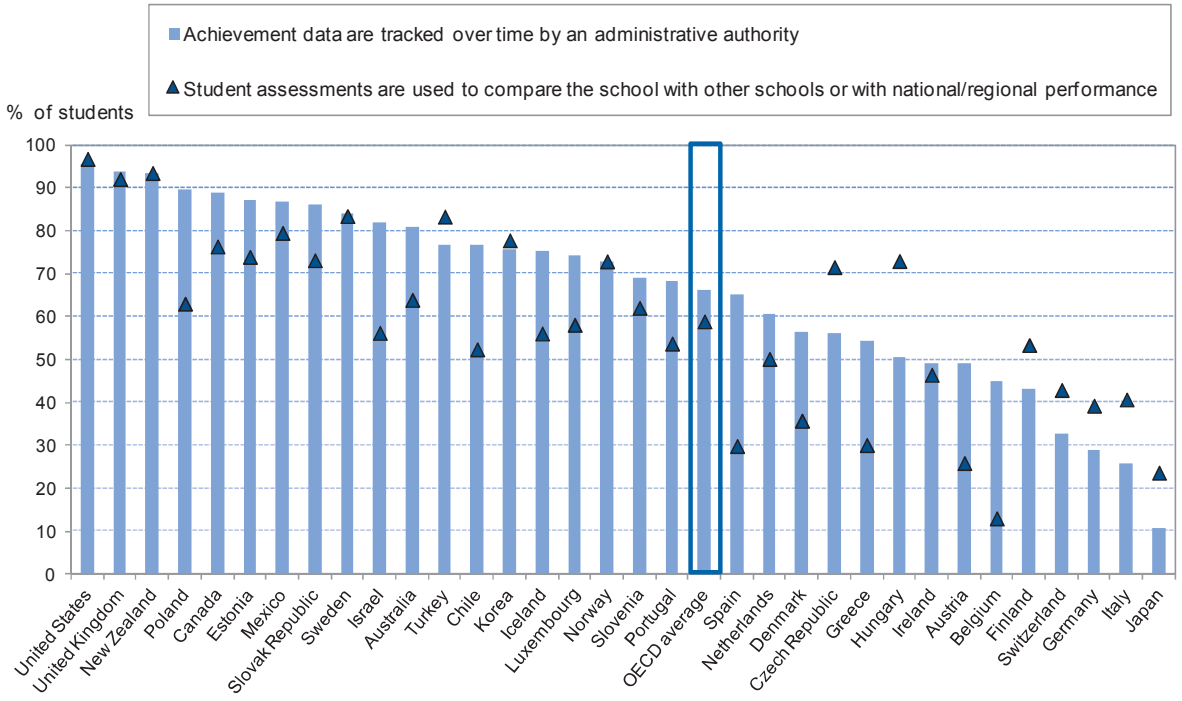
Evidence from PISA indicates that there is an association between the monitoring of schools by an administrative authority and the use of comparative student assessment data within schools (see Figure 8.2). The level of administrative authority is not reported by school principals in the PISA collection, however, it is likely to be local authorities in decentralised systems, e.g. in Denmark, Norway and Sweden. In Sweden and Norway in particular, school principal reports on the tracking over time of student achievement data and the school use of student assessment data for comparative purposes are highly correlated. The data may reflect the variation in those systems of approaches to monitoring at the local authority level. The data also may indicate that more developed monitoring by local authorities is promoting the use of comparative student assessment data at the school level. In Norway, Oslo has a highly developed monitoring system (Nusche et al., 2011b). Schools are divided into eight area-based school groups, each with an area director and about 30 schools. Area directors sit in the Oslo Education Department and will visit schools at least once a year to discuss school results with the school principals. Each area also has its own school support services. Performance management of schools in Oslo incorporates national performance measures and complements these with information gathered in local surveys (both of student skills, e.g. in science and ICT, and stakeholder surveys) and local school inspections in priority areas. There is a clear use of benchmarking and Oslo sets 125 goals overall. Each school is responsible for reporting back on progress against these, but can prioritise particular goals that are most relevant to the school context. Oslo sets six compulsory areas of school improvement/development and Oslo schools are expected to develop a strategic school improvement plan with clear targets.

There are strong correlations between monitoring by authorities and school use of comparative student assessment data also in Australia and Canada where many jurisdictions conceptualise education system evaluation and school evaluation as part of a business planning cycle drawing on the use of comparative data (Figure 8.2). In Australia, all states and territories report annually on the performance of schools in the government sector (public schools) and some may include external school evaluations (see Chapter 6).

Research in the United States has identified the importance of alignment of objectives set by administrative authorities and by schools and, importantly, of a respectful and supportive relationship among schools and administrative authorities. In research on stakeholder perceptions of key characteristics of effective accountability systems and the use of data in schools, Englert et al. (2007) find that stakeholders in schools where performance is improving report more positive perceptions on the use of data to inform school policy and practice. There is also a clear finding that there is a high degree of consistency in reported perceptions among teachers, school principals and the school

administrative authority (superintendents) in improving schools. Levin and Datnow (2012) conducted research in 12 schools in which the use of data was identified as a potential factor in improved student outcomes. This revealed the complex and interrelated nature of school principal actions with the actions of the administrative authority (district leaders) and teachers and students in effectively using data. In general there is consistency in the actions between school principals and administrative authorities, plus between the school principals and teachers. However, there is a looser connection between school principals and students, indicating more autonomy for teachers in this area. The research highlights the key role for school principals in mediating between the administrative authority and teachers and points to the importance of ensuring that the school works towards both goals set by administrative authorities (these are explicit and measureable) and specific goals set for the school. Importantly, the research identified the importance of mutual trust and respect among teachers, school principals and the administrative authorities.

Figure 8.2 School monitoring by administrative authorities (PISA 2009)



StatLink http://dx.doi.org/10.1787/888932791590

Note: Percentage of 15 year-old students in schools where the principal reported this use of achievement data or student assessments. Data are shown for OECD countries.

Source: OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)*, PISA, OECD Publishing, Paris.

However, it must be recognised that the use of school improvement planning does place significant demands on school time and resources and is most effective when all members of the school are engaged in the process and the process is "owned" by the school (see Chapter 6). This necessitates a high level of capacity for self-evaluation

among school principals, teachers, students and increasingly the wider school community. Further, schools with effective school improvement planning establish clear procedures and can benefit from well-developed external support systems.

Reporting and use of results

Methods and approaches to disseminate system level information

There may be a number of reporting requirements at the system level (national or sub-national). These could include the production of an annual statistical report, governmental report on the state of education, reports by specific evaluation agencies, summary reports on results from national assessments, national audit reports on the education sector or part of it, and various reports by sub-national educational jurisdictions (see Chapter 6 also). Different combinations of such reporting is found in the OECD Review countries, but typically aims to report on either the quality of schooling (e.g. student learning outcomes with comparisons across schools and sub-national systems or evaluative reports from external school evaluations) or on specific national strategies or policies. Table 8.9 presents an overview of selected national reporting mechanisms, presenting the most common in the first column and the least common in the last column. The most common reporting approach among the OECD Review countries is to produce ad hoc reports on specific themes, although this is not done in the French Community of Belgium, Canada, Estonia, Finland and Luxembourg.

Although Table 8.9 provides a good overview of different forms of reporting for education system evaluation, these, of course, will vary in design and content significantly. Table 8.9 clearly shows that there may be a number of different reporting mechanisms in place. Certainly, many OECD systems publish specific reports on the results in national assessments. These are geared towards informing policy at the national level, steering the national debate on education, contributing towards action plans to improve the education system and highlighting differences in attainment by different student groups (Eurydice, 2009).

Nineteen systems produce an annual analytical report that is an overview summary report on the education system. The annual analytical report would aim to draw on information reported in each of the different reporting areas and present the main challenges and strengths within the education system. This is a reporting mechanism that has been recently introduced in Poland, Austria and also in Germany. In Austria this report is published every three years in two volumes: one on key indicators on context, input, process, output and outcomes; the other on summarising all scientific evidence on important issues for education policy (Specht and Sobanski, 2012). Box 8.11 presents some examples of annual analytical reports and illustrates that such reports may be developed by different bodies: in the Czech Republic, the Ministry of Education, Youth and Sports develops the annual analytical report; in Norway, this responsibility falls to the Directorate for Education and Training that has specific responsibilities in education system evaluation; and in Portugal, the annual analytical report is developed by an independent advisory body, the National Education Council.

Table 8.9 National reporting on the results of education system evaluation (2012)

	Ad hoc reports on specific themes	Annual statistical publication	Annual analytical report	Specific national reports on the results from international student assessments	Specific reports on the results of evaluations in national priority areas	Information Internet portal	Annual summary report based on school evaluations
Australia	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Austria	Yes	Yes	Triennial	Yes	Yes	No	No
Belgium (Fl.)	Yes	Yes	No	Yes	Yes	No	Yes
Belgium (Fr.)	No	No	No	No	No	No	Yes
Canada	No	Yes	Yes	Yes	No	Yes	Yes
Chile	Yes	No	Yes	Yes	Yes	Yes	Yes
Czech Republic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Denmark	Yes	Yes	No	No[1]	No	No	No
Estonia	No	Yes	Yes	No	Yes	Yes	Yes
Finland	No	No	No	No[1]	Yes	No	No
France	Yes	Yes	Yes	Yes	Yes	No	No
Hungary	Yes	Yes	Yes	Yes	Yes	Yes	No
Iceland	Yes	No	No	No[1]	Yes	No	No
Ireland	Yes	Yes	No	Yes	No	Yes	No
Israel	Yes	Yes	Yes	Yes	Yes	Yes	No
Italy	Yes	Yes	Yes	Yes	Yes	Yes	No
Korea	Yes	No	No	No	No	No	No
Luxembourg	No	Yes	No	Yes	No	No	No
Mexico	Yes	No	Yes	No	No	Yes	No
Netherlands	Yes	Yes	Yes	Yes	Yes	Yes	Yes
New Zealand	Yes	Yes	Yes	Yes	No	Yes	Yes
Norway	Yes	No	Yes	No[1]	Yes	Yes	No
Poland	Yes	Yes	Yes	Yes	No	No	No
Portugal	Yes	Yes	Yes	No[2]	Yes	Yes	Yes
Slovak Republic	Yes	No	No	Yes	Yes	No	Yes
Slovenia	Yes	No	No	Yes	No	No	Yes
Spain	Yes	Yes	Yes	No	No	Yes	No
Sweden	Yes	Yes	Yes	Yes	Yes	Yes	Yes
United Kingdom (Northern Ireland)	Yes	Yes	Yes	Yes	Yes	Yes	Biennial
Total number of systems with this reporting form:	24	20	19	19	18	17	14

Notes: (1) Although there is no specific national report, there is a Nordic regional report analysing results from PISA surveys (*Northern Lights*). (2) However, there has been regular reporting on OECD PISA results.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

In addition to an annual analytical report, countries may also use different ways to heighten the accessibility and use of education evaluation results, for example by providing access to these different reports and education statistics via one central website. In New Zealand, the performance of the education system is assessed against a national education indicators framework relating to six priority areas (Nusche et al., 2012). These indicators are reported in the annual publications The State of Education in New Zealand and New Zealand Schools / Ngā Kura o Aotearoa. More comprehensive and regularly updated information on performance in each of the six indicator areas is also available on line at the Education Counts website (www.educationcounts.govt.nz). The Education Counts website has been developed as a "one-stop-shop" for information on education statistics and research. The website is open to all audiences and brings together a wide range of information including demographic and contextual information; statistical information on educational participation and performance; analyses of education information; and research publications. Norway also uses a common reporting framework for its annual analytical report and online school portal (see Box 8.11).

In Luxembourg, information on outcomes of the education system is presented in a series of different reports on the Ministry of Education and Training's (MENFP) website (Shewbridge et al., 2012). Most reports produced by the Statistics Service within the MENFP include links to electronic data files for readers to download and try to provide readers with data from earlier years for comparative purposes. Since 2002, there has been an annual report on key figures in Luxembourg. This includes information on qualifications and certificates awarded in secondary education. Since 2003/04, a series of specific reports on the fundamental and secondary sectors provide quite detailed statistics on enrolments and certificates awarded. Since 2011, a specific report on the national standardised assessments is published, providing an overview of major results and some examples of the nature of the tests that students sat.

In Portugal, in addition to a series of statistical reports, regularly updated statistics on education are available on line at the Statistics on Education web portal (*http://estatisticas.gepe.min-edu.pt*). It has been developed as a "one-stop-shop" for information on education statistics. The web portal is open to all audiences and brings together a wide range of information including demographic and contextual information; and statistical information on educational participation and completion (Santiago et al., 2012a).

The results of external school evaluations are a rich source of information for education system evaluation. Most external evaluation bodies are obliged to produce an annual report presenting a summary of the major findings in external school evaluations conducted over the previous year. This aims to present an overview of the current state of affairs in the education system and is often called "The state of education". However, there are different approaches to reporting the results of external school evaluations and there may be more in-depth reports on particular aspects of the education system ("thematic reports"). Box 8.11 presents some examples of these different approaches. In the Czech Republic, Portugal and the Slovak Republic, an annual report is published by the external school evaluation body (school inspectorate) summarising the overall findings on the current state of affairs in the education system. In Portugal the 2009-2010 annual report also provided analysis of school self-evaluation activities, as this is a particular policy focus. In the Czech Republic, the annual report includes information gathered as part of both regular external school evaluations and thematic external school evaluations that focus on certain specific aspects of the education system. However, the findings from thematic external school evaluations are also published in specific thematic reports, e.g. on the quality of ICT in basic schools, foreign language education, safety and

health procedures in schools, the development of school education programmes, conditions for admission into secondary schools, and the graduation process in secondary schools. In Northern Ireland in the United Kingdom, the Education and Training Inspectorate publishes a biennial summative report on the education system drawing on evidence from schools that have been externally evaluated over the previous two year period (Department of Education, forthcoming). The most recent report (published in October 2012) presents an assessment of the education system in three areas: achieving value; learning skills; and transforming communities.

Box 8.11 Annual analytical reports on the education system: Czech Republic, Norway and Portugal

In Norway, the major vehicles for reporting results from the national monitoring system are the Directorate for Education and Training's annual analytical report on education in Norway (the Education Mirror) and the web-based School Portal (*Skoleporten*). Both respect a common structure: learning outcomes; learning environment; completion rates in upper secondary education; resources; and school facts. Each edition of the Education Mirror presents a different selection of results in each area depending on the analytical interest and also includes both a special introductory chapter providing examples of schools participating in national initiatives and a final chapter on "Quality development" providing information on national research and initiatives to promote better local monitoring of quality. Results may be augmented by periodic national survey results, but the Education Mirror presents a stable set of national measures, including overall achievement marks and examination results in Years 10 and 12, the results of national assessments and the national student survey and the results from periodic international assessments. The Education Mirror provides an analysis of differences between overall achievement marks and examination results in different subjects. More detailed results at the school level are presented in a password-protected part of the School Portal.

In the Czech Republic, the Ministry of Education, Youth and Sports (MEYS) produces an annual report on its evaluation of the overall education system (the Status Report on the Development of the Education System in the Czech Republic) (Santiago et al., 2012b). The report draws on a set of indicators specifically designed to assess progress towards the long-term policy objectives of the Czech Republic. The indicator set comprises a stable set of basic statistics that are complemented with ad hoc surveys in priority areas. The report sums up the main organisational and legislative changes that occurred in the given year and presents statistical indicators describing the situation and development in pre-primary, basic, secondary, and tertiary education. The report contains information about educational staff in the system, the funding of schools and the labour market situation of school leavers. These data constitute a basis for the development of education policies. This report also typically includes an area of specific focus (e.g. in 2007 and 2008, the implementation of the curricular reform). Individual regions within the Czech Republic also produce their own Status Report in Education to assess progress towards their own long-term policy objectives.

In Portugal, the National Education Council (CNE) comprises a wide range of education stakeholders and provides independent advice to the government of national education issues (Santiago et al., 2012a). The CNE publishes reports on a range of themes such as lifelong learning in the national debate on education, indicators of the education system and the motivation of Portuguese youth for training in science and technology. An important contribution is the annual publication of the report The State of Education, which provides an analysis of key data on the education system. The first issue, The State of Education 2010 – School Paths, offered a more detailed investigation of student pathways in the education system and the second issue, The State of Education 2011 – The Qualifications of the Portuguese Population, provided an in-depth examination of the current qualifications of the population. The report also offers advice on how to improve the quality of basic and secondary education. The CNE makes recommendations for policy development and comments on policy initiatives. In 2011 these covered areas such as school evaluation, the funding of public schools, education for children aged 3 years and under, the reorganisation of the school network and specific education programmes.

Sources: Norwegian Directorate for Education and Training (2007, 2008, 2009 and 2010); Santiago et al. (2012a, 2012b).

In Sweden, the Swedish Schools Inspectorate does not publish an annual summative report (Nusche et al., 2011a). However, an overview of the results of external school evaluations and specifically the thematic quality evaluations is included in the major summative report on the Swedish education system (the National Agency for Education's annual report to the Swedish Government with its assessment on whether the education system achieves its objectives). The annual summative report is taken seriously throughout the system and feeds into the political debate. Results and analysis often feed into policy for school improvement. In addition to the annual analytical report, the National Agency for Education (NAE) also produces regular reports to the Swedish Government on different topics of political relevance (Nusche et al., 2011a). The NAE publishes a series of analytical reports including trends over the last ten years, an annual analytical report on themes of current policy relevance or with a more in-depth examination of factors underlying student performance. These analytical reports draw on results from the monitoring system, plus other reports or reviews that have been conducted. Such reports include suggestions for the Government, local authorities and schools.

The participation in international work on education indicators and in international studies also provides key information for education system evaluation. Nineteen systems publish specific national reports on the results of international student assessments (Table 8.9). Such reporting allows a more detailed discussion of the results for that system benchmarked against international performance. Canada presents an example of an education indicators report aligned to international indicators as reported in the OECD's *Education at a Glance* (Box 8.12). Typically, countries draw on the results of international indicator work and international studies when compiling their annual analytical report.

In Austria, international evidence currently forms a substantial part of the triennial analytical report on the education system – in particular information on student outcomes, although the new national assessments will complement this information in future editions. In the Netherlands, the annual education indicators report (most recent edition is Key Figures 2007-2011 Education, Culture and Science) includes a special chapter on Dutch education in an international perspective. This incorporates information from international indicator collections and international student assessments. An innovation in the most recent edition is to also benchmark against the top five performing systems on each international indicator. This indicator report provides "a quantitative picture and a brief descriptive analysis of the developments in the areas of education, culture and science" (Dutch Ministry of Education, Culture and Science, 2012).

Box 8.12 Examples of reports on the education system

Summary report based on school evaluations: the overall evaluation of the education system

In the Netherlands, the Inspectorate of Education must produce an annual report on the State of Education in the Netherlands (Departmental Arrangement on School Inspection, April, 2003). This typically comprises two major sections: general introductory chapter on the State of Education and the Supervision of the Inspection; and description of the different educational sectors. On the basis of the inspection framework, the Inspectorate judges the proportion of schools that is above or below certain achievement standards and can identify strengths and weaknesses within the education system. In the 2011 report, a third section addresses the following themes: educational attainment, the quality of educational governance, finance and legitimate spending of funds, teacher quality and the quality of education for students who need special care. The report includes evidence from national and international sample assessments and other research. However, the report provides a "unique evaluative contribution" with its descriptions and evaluations of classroom teaching and school functioning resulting from the systematic school inspection, based on the inspection framework, standards and indicators (Scheerens et al., 2012).

Summary report based on school evaluations: evaluation of specific national priority areas

In New Zealand, the Education Review Office (ERO) integrates national evaluation topics into its individual school reviews (approximately 600 primary and secondary schools each year). These always include the success for Māori and Pacific students and other topics are decided in collaboration with the Ministry of Education and other government agencies. Such topics of system-wide evaluation reflect national priorities and inform the development and implementation of education policy and practice. ERO reports on national evaluations of education sector performance in anything between 12 and 20 specific reports each year. A second reporting format is "Good Practice Reports". These typically use a case study approach to identify the nature of effective practice in schools (Nusche et al., 2012).

Annual indicator report

In 1996 the Canadian Education Statistics Council (a partnership between Statistics Canada and the Council of Ministers of Education, Canada [CMEC]) published the first Canadian education statistical indicators report (Fournier and Mildon, forthcoming). The report *Education Indicators in Canada: Report of the Pan-Canadian Education Indicators Program* was also published in 1999, 2003, 2005, 2007, and 2008. These reports provided "a set of statistical measures on education systems in Canada for policy makers, practitioners, and the general public to monitor the performance of education systems across jurisdictions and over time." Since 2008, the Canadian Education Statistics Council publishes an annual indicator report based on the OECD's *Education at a Glance* (Education Indicators in Canada: An International Perspective).

Sources: Scheerens et al. (2012); Nusche et al. (2012); Fournier and Mildon (forthcoming).

Intended users of system evaluation results

The results of education system evaluation serve the purpose of accountability of educational policies, but also provide useful information for the further development of policies and systems. Ministries draw heavily on the results of education system evaluations, but these can also feed into the work of stakeholders throughout the system. Box 8.13 shows an illustration of the intended use of system evaluation results in the Slovak Republic.

Box 8.13 An illustration of the intended use of the results of education system evaluation: Slovak Republic

In the Slovak Republic, the results of education system evaluation can have the following uses by different stakeholders:

- Ministry of Education: results provide background for accepting and forming legal measures (e.g. amendments to regulations) and other measures (e.g. changes in the content of national school programmes, incentives for continual education, methodological and organisational guidelines, publishing textbooks, workbooks, teaching materials etc.).
- National Institute for Certified Educational Measurements: informs ways to increase the objectivity of external measurements (national assessments in ISCED 2 [*Testovanie* 9], student examinations at ISCED 3 [*Maturita]*), to create self-evaluation models, to develop and adjust quality criteria and indicators.
- Slovak State Schools Inspectorate: to verify new methodologies for external school evaluation and develop and refine criteria and indicators for external school evaluation.
- Regional school authorities: to publish methodological and organisational guidelines for school leaders, guidance for school founders, etc.
- Primary and secondary school founders: as a background for the evaluation of school leaders and to create better conditions for schooling processes that aim to improve student outcomes.
- Primary and secondary schools: for primary schools to improve student outcomes in order to improve transition of students into secondary schools; for secondary schools to improve student outcomes in order to improve students' further progress in education or successful insertion to the labour market.

Source: Hajdúková et al. (forthcoming).

The influence of national political and legal contexts on system reporting

The broader political and legal framework has implications for the reporting of system level information. Freedom of Information laws mean that centrally collected information – including student assessment results – can be requested by any citizen. In France and the Netherlands, the media have been instrumental in compelling the government to publish student examination results at the school level (Rosenkvist, 2010). In England in the United Kingdom and Australia, the media publish school rankings or "performance tables" drawing on officially published data. The Australian official school performance website (My School) operates a security system to prevent the media being able to download officially published data to create school ranking or performance tables, as this practice is not condoned by the Australian government. Official policies aim to prevent the simplistic publication of school performance measures by the media that can mislead the public due to a lack of adequate contextual information to meaningfully interpret the results (see also Chapter 6). The OECD Review has also revealed examples of non-official websites compiling and presenting school performance data in Ontario and the Atlantic Provinces of Canada and in the Slovak Republic. The evidence on the effect of publishing student examination or assessment results in school performance tables is mixed, with some studies showing a positive relationship with student performance

results, but others showing unintended strategic behaviour by schools, teachers and parents (Rosenkvist, 2010).

In Austria, Belgium, Luxembourg and Slovenia, official documents state clearly that national assessments cannot be used to rank schools and in Finland despite media pressure, the national consensus in the ensuing debate was against publication of results (Eurydice, 2009). In Denmark, the national assessments were designed for the primary purpose of providing high-quality diagnostic information to schools and teachers for improvement of student learning and school development. As such, a specific law specifies that only a national average will be published, but that all other data will be reported comprehensively to schools and local authorities, but not made available to the public. In Northern Ireland in the United Kingdom, national assessments are administered purely for diagnostic use by schools and teachers (see Table 8.3). The results are not collected centrally, so none of this information feeds into system evaluation. However, the results of teacher marked student assessments in the cross-curricular skills of Communication, Using Mathematics and Using ICT are collected centrally at three key stages in compulsory schooling and used as performance indicators in education system evaluation.

Communicating education system evaluation results clearly and comprehensively

An objective of system level evaluation is to inform the general public and the range of stakeholders on goal achievement in the education system and the impact of national strategies and policy initiatives. A need in system evaluation is to design a comprehensive presentation of results of system evaluation for stakeholder use. While countries often collect large amounts of data and statistics at the system level, there is frequently significant untapped potential for integrating and using the available data. This is sometimes the result of insufficient consultation between interested stakeholders and agencies on how to best manage and present data for optimal use by different audiences. The challenges are: how to best organise the collection and analysis of key information at the national level; to clearly communicate results of system evaluation; and to ensure the effective use of results by stakeholders throughout the system.

In communicating education system evaluation results, it is important to bring together important facets of reporting principles. Based on international best practice, the OECD defines the quality of its own statistical activities as going beyond the accuracy of the data to ensuring its "fitness of use" in terms of user needs (OECD, 2012). Thus, statistical quality depends on accuracy and coherence, as well as on:

- Relevance: This relates to the identification of user groups and their needs. User groups may change over time and their needs for data may evolve. This highlights the importance of having processes in place to determine the views of users and the uses they make of the data.
- Credibility: This refers to trust in the objectivity of the data. This implies that data are perceived to be produced professionally in accordance with appropriate statistical standards with transparent policies and practices for their reporting and release, e.g. the release is not timed in response to political pressure.
- Timeliness: This refers to the length of time between the availability of data and the event or phenomenon they describe, but also to the punctuality and clarity of reporting schedules.

- Accessibility: This includes the suitability of the form in which the data are available, the media of dissemination and the availability of metadata.
- Interpretability: This reflects the ease with which the user may understand and properly use and analyse data. This relates to definitions of concepts, terminology and information describing the limitations of the data.

Box 8.14 presents an example of a strategy to address these key principles.

Box 8.14 Strategy to make education system evaluation results more accessible in Sweden

The National Agency for Education (NAE) in Sweden has a clear commitment to make key results at the national level more accessible. All official statistics for the Swedish education system are reported on the NAE's website. Both full reports and statistical tables (in MS Excel format) are available. In 2001 the NAE redesigned its reporting website to present more clearly statistics at the national, local authority and school levels (including analytical tools for comparing school performance [SIRIS database and SALSA analysis tool]). For example, up until 2000 results were published in four separate papers: a summary of main results plus analytical papers by researchers on results in English, mathematics and Swedish and Swedish as a second language.

As well as a clear presentation of national assessment results, the annual report on the national assessments includes content analysis of national assessment results for each subject by different researchers (e.g. Stockholm University on mathematics, Göteborg University on English, Uppsala University on Swedish and Swedish as a second language). The content of the reports may vary and highlight different areas of interest, e.g. the 2009 report included trend results from 1998 to 2008.

Further, the NEA publishes a transparent reporting schedule, favouring the publication of reports in different stages to allow a more timely feedback of results.

Source: Nusche et al. (2011a).

There is also a need to strengthen the channels to feed results from the national monitoring system back into policy and practice. The OECD Review has identified a varied suite of reporting used by countries (Table 8.9). A lack of clear reporting and dissemination runs the risk of a more negative impact of system results. For example, a study on the impact of PISA results suggests that the media play the most important role in countries without comprehensive strategies for the dissemination of PISA results, which is likely to have a negative impact (Hopkins et al., 2008). Slovenia presents an illustration of a more proactive strategy to disseminate results of international comparative studies (Box 8.15). One way to increase the immediate impact of results is to organise conferences to bring together stakeholders to learn about and debate these (Box 8.16). It is worthy of note that most countries in which an OECD Review has been conducted have also organised specific dissemination events to discuss the review findings and/or have included these in wider educational conferences.

Box 8.15 Promoting the use of results from international comparative studies in Slovenia

In Slovenia, system evaluation results have greatly improved the overall level of discussions on educational policy and practice. Specific national reports on results from international comparative studies are produced (e.g. OECD's PISA and TALIS, IEA's TIMSS, PIRLS and ICCS). The release of the national report and international results is typically accompanied by a press conference and other dissemination activities to increase the impact of the results. System evaluation results generally receive much media attention and are closely followed by different stakeholders, including teacher unions, teacher associations, civil society initiatives, parent associations and others. Results are also packaged for different target groups, e.g. for mathematics teachers when results are of particular relevance to the further improvement of teaching practice in mathematics. Teacher training seminars and other dissemination activities promote the wider use of international assessment results. This also leads into further policy development and the introduction of various policy or practical initiatives.

Source: Brejc, Sardoč and Zupanc (2011).

Box 8.16 Stakeholder discussion of major education system results in Denmark and the Flemish Community of Belgium

Belgium (Flemish Community)

The National Assessment Programme is conducted on an annual basis. In addition to reporting on the major results and analysis of factors associated with achievement (on the Ministry's website, in a specific brochure and via a colloquium), the Flemish Ministry of Education and Training promotes the discussion and use of results among stakeholders. Since the inception of the National Assessment Programme the Ministry has always sought feedback on the results from key stakeholders, but has done so more actively since 2010, seeking feedback from for example, school support bodies (School Advisory Services), the Flemish Inspectorate of Education, Institutes responsible for initial teacher education, researchers and publishers etc. After these consultations, the Ministry organises an open conference to discuss possible actions to promote and improve school quality.

The final step is that the Ministry of Education and Training and other stakeholders engage in concrete actions based on the results of the National Assessment Programme and the discussion with stakeholders. Possible improvements include: updating of the attainment targets; developing or adjusting curricula or teaching materials; adjusting initial teacher education and/or teacher professional development; adjusting school policies; introducing new initiatives to support specific student groups.

Denmark

In Denmark, the School Council for Evaluation and Quality Development of Primary and Lower Secondary Education is an advisory body with representatives from all the major stakeholder groups. The School Council holds an annual meeting at which major evidence on the education system is discussed and debated. This is also supported by the production of an annual report presenting a summary of major research and evidence.

Sources: Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010); Shewbridge et al. (2011).

In Mexico, the National Institute for Educational Assessment and Evaluation (INEE) organises large conferences bringing together on average about 500 individuals including representatives of state authorities, supervisors, heads of sector, civil society organisations and education specialists (Santiago et al., 2012c). It is also engaged in capacity building with state evaluation authorities with a view to strengthening the technical, statistical and analytical skills of evaluation teams at the state level. INEE has established strong relationships with the media and aims to inform the national education debate through press releases, interviews and courses for journalists on the interpretation of education data (INEE, 2006b).

Making better use of the results of system evaluation in planning and policy development

With the collection of new measures of the education system, much information is generated. Some systems are challenged to exploit this adequately. There may be considerable challenges in ensuring that system level data are used to their full potential in analysis which could be useful to inform policy development. The use of system level information is often limited in terms of research and empirical analysis. Brokerage agencies were identified in an OECD project as an effective way to bring together disparate research communities and policy makers (OECD, 2007b). These take different forms in different OECD countries, but in all cases they aim to encourage dialogue between policy makers, researchers and to provide tools and capacity-building to evaluate what works in education. However, the project identified persisting challenges, including addressing the tension between the time required for solid research and the necessity of quick results for policy making and the adequate dissemination of findings to all stakeholders, including media, parents and students.

In Sweden, the Ministry of Education and Research has a policy to improve links between the research society and policy by, for example, collecting reviews of research on different thematic areas (Nusche et al., 2011a). Since 2008, the National Agency for Education has had the task of disseminating research results. In 2009 an official government evaluation on the use of evaluation results in the Swedish system suggested that the NAE conduct systematic overviews of research and share the results in an easily accessible format with schools (SOU, 2009).

Box 8.17 presents examples of how some systems have implemented mechanisms aiming to more effectively build a core evidence base and to package this in a more digestible form for policy makers and other stakeholders.

Box 8.17 Support for evidence-based policy making within ministries

In the Netherlands, the Ministry of Education has established a "Knowledge Directorate" to support an evidence-based approach to policy development. It acts as a clearing house of scientific research. The decision to establish the Knowledge Directorate reflects the importance attributed to evidence-based policy making in the Netherlands and aimed to mitigate three identified barriers to the effective sharing and use of evidence in policy making (OECD, 2007b): the increasing volume of knowledge and information makes it difficult to identify relevant research, to interpret this correctly and to link it with existing evidence; knowledge tended to be compartmentalised between and within departments in the Ministry; and senior government officials may lack the understanding to examine adequately the evidence base of policy proposals.

In New Zealand, the Ministry of Education runs an Iterative Best Evidence Synthesis programme to compile "trustworthy evidence about what works and what makes a bigger difference in education". A Strategy and System Performance Group within the Ministry has core responsibility for system evaluation and assessment and runs this programme. Evidence collected in this programme showing impact on student outcomes feeds into the development of education indicators that are used to evaluate the performance of the education system overall and the quality of education provided in individual schools. The policy significance of the Best Evidence Syntheses has been recognised by the International Academy of Education and the International Bureau of Education. Summaries of recent Best Evidence Syntheses are published on the UNESCO website, see *www.ibe.unesco.org/en/services/publications/educational-practices.html*.

See also: *www.educationcounts.govt.nz/topics/BES*.

In Northern Ireland in the United Kingdom, the Northern Ireland Statistics and Research Agency (NISRA) has progressively organised its structure and services around different major users and demands for education statistics. For example, NISRA includes a specific section that lends support to the Education and Training Inspectorate. NISRA provides regular focused briefings for policy makers and identifies developing trends. In addition to increased requests from policy colleagues to provide an evidence base for the development and monitoring of education policies, NISRA answers requests for data from researchers, members of the public, the media and the Northern Ireland Assembly. NISRA provides a number of statistical press releases. These comply with the United Kingdom Statistics Authority Code of Practice, which specifies a number of reporting guidelines, including that statistics are well explained and readily accessible. As such, each statistical release by NISRA includes sufficient commentary to enable users to meaningfully interpret the information. These usually take the form of a few introductory lines, major bullet points of key results and graphics showing trends and then the full set of results in tables. See for example:
www.deni.gov.uk/year_12_and_year_14_examination_performance_at_post_primary_schools_201112__2_.pdf.

Sources: OECD (2007b); New Zealand Ministry of Education (2010); Shewbridge et al. (forthcoming).

Making system level data useful for managing school sub-systems

Another area in which efforts are needed is to ensure schools and local education authorities are provided with useful information for their own management. While schools report the data for the national information systems, they sometimes do not receive a statistical analysis of their profile from national authorities to support them in their internal analysis and further planning. It would be important for schools to compare their own data with indicators aggregated to meaningful benchmark groupings (e.g. the

local authority or regional level, schools with similar student composition, schools with similar pedagogical philosophy, etc.).

Local authorities systematically receive feedback of student results in national assessments in about half the European systems (Eurydice, 2009). In Scotland (United Kingdom), local authorities are offered the possibility to increase the sample size in their jurisdiction as part of the national sample survey. This provides richer feedback for local system management. Similarly, in the Flemish Community of Belgium, schools are offered the chance to administer the national assessment, even if they are not selected in the national sample. Schools opting to take the national assessment receive a detailed feedback report on student performance in relation to national averages (see also Chapter 6). Several countries opt to over sample in OECD's PISA, so as to provide benchmark data on performance of sub-systems and these results are published in annexes to official OECD reports (e.g. OECD, 2010a).

System level results can feed into local monitoring systems and ensure better alignment between national and sub-national goals. In Australia, all state and territory government departments produce an annual report on major activities, including both financial and performance information (Santiago et al., 2011). A common feature in the 2009/10 government reports is the prominence of results from the Australian National Assessment Program – Literacy and Numeracy (NAPLAN) in the performance monitoring. The exact format for reporting NAPLAN results varies according to the emphasis on different monitoring goals in each state and territory. The majority of jurisdictions report according to the national minimum standard. In Sweden, the Swedish Association of Local Authorities and Regions (SALAR) publishes results of national assessments with performance ranking for local authorities and schools and attempts to identify effective strategies and practices (Nusche et al., 2011a).

The OECD Review has revealed an increasing importance of information systems to both collect, compile and manage information and also to disseminate the results of education system evaluation throughout the system. Information systems can automate reporting tasks, thus avoiding duplication of efforts and improving cost-efficiency, but can also be used to inform financial decision making (OECD, 2010b). Information Internet portals are used in 17 of the systems participating in the OECD Review (Table 8.9). In Norway, the Directorate for Education and Training manages a Compulsory School Information System (GSI) which contains basic descriptive statistics on school demographics, resources and organisation. Such information systems are a typical feature in OECD countries and are being developed where these do not already exist, e.g. in the Slovak Republic (Shewbridge et al., 2013). In other cases, there may be information systems that are not specifically introduced for education system evaluation, but that provide key information that can be used to this end. In Korea, the Educational Information Disclosure System was launched in 2000 and includes school performance data. As such, this lays the foundation for the comparison and analysis of school systems and policies, but as yet is not used to evaluate the effectiveness of the education system as a whole (Kim et al., 2010).

Information systems are often developed at the sub-national level for use in evaluating local school systems (see Box 8.18). Further, such systems generally aim to promote the use of the results of different tools used in education system evaluation in schools for their own self-evaluation (see Chapter 6). In Australia, several state and territory governments have developed specific analytical information systems to optimise school use of results from the national full-cohort assessments, as well as from assessments run by local governments (see Chapter 3).

Box 8.18 Information systems for evaluation of local school systems: The United States

A study by the U.S. Department of Education (2010) that ran during 2006-2008 examined how education data systems varied across educational districts and how they were used to aid decision making. The report uses the Wayman (2005) classification of four types of electronic student data systems:

1) Sudent information systems providing real-time access to student data on attendance, enrolment, marks and schedules;
2) Data warehouses providing access to current and historical data on students, finances and staffing;
3) Instructional or curriculum management systems providing planning tools, links to state content or performance standards and communication tools; and
4) Assessment systems supporting the organisation and assessment of benchmark data.

In general there was a huge increase in reported availability of data systems. Virtually all school districts had student information systems storing basic information on enrolments and attendance and 79% reported having an assessment system to organise and analyse benchmark assessment data. The least common system was on instructional or curriculum management (64% of school districts). The major challenge reported by school districts was to link these multiple data systems to better support decision making and in particular to better link student data to instructional practice.

The report found that most systems had developed in response to accountability requirements and less than half the school districts could link outcomes to processes in order to monitor and promote continuous improvement. An example here is that only 42% of school districts could link student performance to participation in particular programmes. The most common school district policies to promote schools to use data was to incorporate this in school improvement planning, providing professional development activities and support positions for system implementation and developing data generation and analysis tools. Examples of support provided by school districts included: technical expertise to schools, "data coaches" available to schools, creating easy-to-read data "dashboards" to make information more accessible to teachers, and developing benchmark and formative assessments providing teachers with more timely data on student progress.

Source: U.S. Department of Education (2010).

In general, the use of information systems for accountability is well established. However, their potential use to inform improvement in school systems is often under exploited. The OECD has an on-going project bringing together experts and policy makers to examine how to best design information systems to inform educational innovation and improvement. Preliminary findings suggest many such systems have been developed relatively recently and could be improved by: improving the speed of feedback to teachers and schools; better integrating data systems and learning management systems; including information on generic and higher order learning outcomes; and developing comparison tools especially at the student level (OECD, 2010b). In an overview of school performance feedback systems provided in Australia, the Netherlands, the United Kingdom and the United States, Visscher and Coe (2003) identify the following features as being likely to lead to greater school use: flexible systems that can meet varying information needs among schools; systems with valid information with good coverage of school quality, allowing more in-depth analysis; and the systems are accompanied by comprehensive, tailored

reform and support strategies. The use of information systems is well established in many states and school districts of the United States. A recent study sheds some light into how school districts may make better use of these for improvement (Box 8.18).

Pointers for future policy development

This chapter has reviewed the approaches countries are taking to education system evaluation in light of available research and evidence. The policy suggestions that follow are drawn from the experiences reported in the Country Background Reports, the analyses of external review teams in Country Reviews and the available research literature. It should be stressed that there is no single model or global best practice of education system evaluation. The development of practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing policies for school evaluation for different countries.

Governance

Ensure a broad concept of education system evaluation within the evaluation and assessment framework

An initial priority is to ensure a broad concept of system evaluation as the wide range of system level information which permits a good understanding of how well student learning objectives are being achieved. System evaluation should include a varied set of components such as broad measures of student outcomes; system-level indicators with basic demographic, administrative and contextual information; information systems; and research and analysis to inform planning, intervention and policy development. A strategic approach to system level evaluation would benefit from clear national objectives and priorities so progress against these can be assessed. System level evaluation should include the production of an annual report with an assessment of whether or not the education system is achieving its objectives.

Ensure policy making is informed by high-quality measures, but not driven by their availability

Measures of education system performance must be broad enough to capture the whole range of student learning objectives. However, it is not always possible to devise indicators and measures of good quality across all the objectives of the education system. Hence, it needs to be recognised that policy making at the system level needs to be informed by high-quality data and evidence, but not driven by the availability of such information. Qualitative studies as well as secondary analysis of the available measures and indicators are essential information to take into account in policy development and implementation.

Ways to ensure that indicators and measures are of high quality include: ensuring systematic collection to agreed definitions of existing information at different levels in the system; promoting data quality improvement; undertaking research to shed light on some of the "gaps" where systematic collection is too costly/not feasible; and developing a long-term strategy to improve measurement tools for future information needs.

To ensure that education policy is not driven by the availability of data, there should be a systematic review of the availability of key measures in a meaningful and nationally agreed education system evaluation framework. This should be based on the mapping out of key objectives for the education system, followed by a set of goals or targets to be realised. The next stage is a systematic mapping out of available measures, plus where measures are available a technical note on their validity and/or limitations for interpretation. An analysis of this framework will provide information on key gaps in data availability and also in limitations of existing measures. This will be the foundation of strategies to prioritise further measurement development and/or refinement according to the national political priorities and long-term goals. This mapping out is also a critical exercise in reminding all stakeholders of the full spectrum of national priorities and goals and clearly showing that not all of these are currently measured. A final step is to ensure a qualitative analysis in priority areas for which there are currently no measures and to feed the results of this analysis into the policy making process.

Situate education system evaluation in the broader context of public sector performance requirements

Education system evaluation should be thought of in the broader context of established performance measurement frameworks for the public sector. It is a common expectation for Ministries and specific bodies with a role in evaluation (e.g. external school evaluation bodies or national assessment or evaluation agencies) to publish annual reports to show accountability for their performance. This stimulates the need for procedures in place to regularly monitor progress in the education system or particular aspects of the education system under their responsibility. In systems where high level targets are set by the government related to productivity, competition or general economic and social improvement, this has implications for the education system. There may be a need to monitor the progress towards achieving these high-level targets and where necessary to establish adequate systems to collect evidence on progress.

Increasingly, national audit offices conduct evaluations of different aspects of the education system. An often cited finding is the lack of evaluation of particular education policies or programmes. In some systems this may be due to the lack of appropriate outcome data. However, there may be a more fundamental need to establish a culture of evaluation. Some systems have aimed to stimulate this by attaching requirements for schools to monitor and evaluate the implementation of particular education programmes as a funding condition. This can also be a way to stimulate school self-evaluation. The evaluation outcomes can be used to monitor effectiveness at the education system level. For major reform initiatives, there should be official evaluations of implementation, including adequate consultation and feedback from key stakeholders. Evaluation results should feed back into the further implementation of reforms and where necessary the design and offer of related support structures.

Procedures

Develop a national education indicator framework

The development of an education system framework allows the systematic mapping of available information against education system goals (see above). A major support for this is the development of an education indicator framework. This should provide a stable set of key indicators on the education system to allow the monitoring of changes over

time. But the indicator framework may be augmented periodically to reflect areas of national priority. Aligning the national indicator framework with international indicator frameworks provides the benefit of streamlining data collection and providing international benchmarks for the national education system.

The use of a stable indicator framework also promotes the regular collection of information from schools and other stakeholders. This allows the possibility to develop more efficient data collection systems to both ease the reporting burden on schools and to ensure more efficient compilation of key information at the national level. For example, key information may be collected systematically on an annual basis, but other information can be requested on a cyclical basis. This also provides the basis to ensure the feedback to schools of useful benchmarking information.

Design a national strategy to monitor student learning standards

Student assessment provides keystone indicators for assessing system performance. Assessments of student learning provide evidence by which policy makers, the public, administrators, educators and parents at the national and local levels can gauge both students' current performance relative to student learning objectives and the extent to which improvement goals are being realised. To this end, designing a national strategy to monitor student learning standards provides a strong basis for education system evaluation. This should promote the combined use of a variety of tools to collect information on a broad set of student outcomes.

Many education systems use the results of student summative assessment, such as examinations and final marks at the end of compulsory education, to monitor student learning outcomes. However, such measures do not allow for comparison over time, as each year the content of the examinations changes. This renders them of limited use in assessing the progress in learning outcomes against national standards. It is important, therefore, to provide other measures to complement these and to allow the monitoring of trends (see below). In addition to the possibility of participating in international student assessment surveys, education systems may want to design national student assessments with the primary purpose of providing information to monitor key outcomes of the education system. In contrast to international assessments, these national assessments would be aligned with national learning objectives and can be administered more regularly or on a more flexible schedule to provide more timely feedback on progress (international surveys generally run on a 3-to-5-year cycle). There may be, for example, the need to monitor the implementation of a new national curriculum.

Education system performance monitoring based on national student assessment programmes can take a variety of forms. Periodic sample-based student assessments/surveys can allow greater breadth of measurement, fuller coverage of the curriculum and avoid distortions deriving from "teaching to the test". They can be carried out at comparatively low cost. In turn, full-cohort student assessments have the advantage of potentially providing feedback to schools on classes/students, but are narrower measures that cannot realistically include a full coverage of the curriculum.

Large-scale standardised assessments are often limited to written formats such as multiple-choice questions or questions requiring a short closed-response answer (e.g. yes/no) that are easiest to score and most cost-efficient to implement: such assessments may only draw upon a limited set of students' skills. National assessments can also include performance like tasks, where students are scored on open-ended questions, such as written essays, oral communication skills, reasoning processes,

collaborative problem solving, and so on. These are often seen as being more effectively aligned with curricula that emphasise development of higher-order thinking skills and capacity to perform complex tasks. However, they require much investment to ensure the comparability of scoring.

Countries often design standardised national student assessments to monitor student progress towards achieving student learning objectives but these tend to cover a small set of curricular areas, typically the language of instruction, mathematics and science. To have reliable measures of performance across broader curricular areas a good option is to introduce sample-based national monitoring surveys. The sample-based surveys test a statistically representative sample of students in specific school years and in a given set of curricula or cross-curricula areas. A possible approach is to test a small number of subject areas each year for given grades in 3- or 4-year cycles with different subject areas every year. Such sample-based surveys would allow the assessment of a broader range of curricula content and cross-curricula skills such as civic and citizenship skills, learning-to-learn skills and problem solving skills and allow benchmarking of different regions or specific student groups on an externally validated measure.

Ensure the collection of qualitative information on the education system

There are different approaches to collecting qualitative information on the education system. This can be done via external school evaluation, specific research studies or by developing and administering stakeholder surveys.

Education systems with an external school evaluation mechanism have the possibility to collect a rich set of evidence on different qualitative aspects of schooling. Such information is collected via the regular external evaluation of individual schools and can be compiled to provide an overview of evidence on the education system. However, there is also the possibility to include the evaluation of priority areas as part of external school evaluations – either integrated in regular evaluation of individual schools or in a special thematic evaluation in a sample of schools. This holds strong potential for policy makers to examine a broader set of evidence against current policy priorities. However, the collection of such thematic evidence needs to be balanced in the system of overall external evaluation of individual schools and implications for the regular external school evaluation cycle should be carefully considered (see Chapter 6).

There is also the possibility to administer qualitative surveys to seek key stakeholder feedback on broader outcomes, such as motivation to learn, subject engagement and different aspects of the school climate. Analysis of the perceptions of students, teachers, school leadership and parents on these different aspects can feed into the policy debate by providing evidence on a broader set of student learning outcomes, as well as help shed light on some of the factors associated with better student learning outcomes. These can also provide useful input on processes and contextual information to help interpret information on outcomes.

There are several options to consider on how best to include stakeholders' perceptions of the teaching and learning environment in the national monitoring system. There could be a national questionnaire administered to a sample of students, parents, school leadership and teachers in the education system to collect views and perspectives about a range of aspects such as attitudes to learning and assessment, perceptions on the implementation of policies, well-being, engagement, satisfaction, etc. Another option is to include a questionnaire to students during the administration of national student assessments. Certainly, the collection of information from students, school leadership and teachers

during the administration of international surveys has led to informed analysis of how different reported factors relate to student performance. The use of student and parental surveys could also be encouraged at the school level through the development of a national template to which schools could add issues more related to their specific circumstances.

Assure the monitoring of changes over time and progress of particular student cohorts

System evaluation needs to place significant emphasis on the monitoring of student progress in contrast to achievement levels at a given point in time. This sends a firm signal on the importance of the improvement function in evaluation and assessment. To achieve that, it needs to undertake the monitoring of both student results over time and the progress of particular student cohorts. This should be a key purpose of specific assessments to monitor education systems, whether full-cohort national assessments or sample-based monitoring surveys. A prerequisite is to ensure the comparability of results over time by keeping a stable element of items in the tests and releasing only a proportion of the items for use by teachers after the assessments have been administered. With a stable difficulty level for each assessment from year to year, national student assessment results provide a useful indicator on changes in student performance over time – one which complements the international trend measures.

Taking a more longitudinal approach to analyse student results could provide additional useful information that allows analysing student pathways. There are several approaches to this, including strategies to monitor student progress through different stages of the education system, designing national monitoring measures to show the progress in measured skill areas of particular student cohorts and the follow-up of students subsequent to their completion of compulsory education.

Education authorities could explore ways to link information for individual student progress through primary and lower secondary as well as upper secondary. This would permit schools to better follow student transitions between schools. Also, the eventual use of a student identifier could encourage more longitudinal studies of student progression and transition to upper secondary education. Further, this would allow analysis of earlier educational pathways in identifying success and risk factors for students in upper secondary education.

Another possibility is to make a strategic use of the national monitoring measures to provide indicators on the progress of particular student cohorts through school education in some curricular areas. With individual student identification numbers, results from the national assessments could be linked across cohorts to report on the success of a given cohort on national assessments in a given number of years. Or the national measures could be designed to measure student progress across key stages of the education system. That is, the tests would be based on a common scale of difficulty across different years, thereby allowing the comparison of a given student cohort in for example Year 1, Year 4 and Year 7 on the same national assessment.

Finally, there could be national surveys to look at how groups of students with different characteristics and academic profiles succeed in later years of education. This could be done by introducing longitudinal surveys to follow students on their pathways beyond secondary education, including in both higher education and the labour market. Further, some education systems have capitalised on linking national longitudinal surveys with cohorts of students tested in international assessments to measure their progress,

pathways and analyse risk factors judged by performance at earlier stages in the education system.

Ensure collection of adequate contextual information to effectively monitor equity

Within the overall evaluation and assessment framework, education system evaluation has arguably the strongest potential to pay attention to equity issues and to inform policies on how to address these and to target support more effectively. The monitoring of student performance across specific groups (e.g. by gender, socio-economic or immigrant/cultural background) as well as across local and regional authorities should receive priority. Analyses from international and national research have proven the strong influence that socio-economic and other contextual factors have on student performance. Therefore, when comparing performance measures across different regions, local authorities and schools, an imperative is to make comparisons meaningful in the light of differing contexts. National research into the how student background characteristics and school contextual characteristics are associated with student performance can identify the type of information that is most pertinent to collect systematically. Typically, information on the student socio-economic background may include a mix of the following factors immigrant/cultural/linguistic background, parental level of education, occupation and income level.

There is a concern when publishing average results of national assessments or examinations at the school level, that these measures are accompanied by adequate information on the school context (see Chapter 6). This is particularly pertinent when education systems use school results to hold schools accountable, for example, schools may be required to meet certain student performance targets and/or demonstrate the progress of students through different stages of education. In such a context, measures of school performance need to reflect the starting point of student performance and adjust for factors which are outside a school's control, but which evidence has proved to have an impact on student performance (e.g. socio-economic background). There is a tension between the use of measures in school accountability systems and the use of measures for comparing performance over time and across the education system. This may require the development of specific measures for the different purposes.

Capacity

Establish and secure capacity for education system evaluation

In some countries, there is a strong case to build central competencies in evaluation for policy purposes. A major signal of political support for evaluation is to establish national bodies competent in this area. This may include specialised research institutes, regional school review boards, national monitoring and quality development agencies, national evaluation institutes, etc. An important aspect of education system evaluation is to ensure the adequate use and interpretation of system evaluation results, including research, results from regular monitoring tools and the evaluation of educational programmes. This necessitates adequate analytical capacity at the national level to fully exploit existing information by ensuring statistical, analytical and research competencies. The results of such analysis and research should be systematically reported in major publications such as annual reports on the state of the education system. This should in turn set priorities for further research and analysis and resources should be invested accordingly. The active participation in international evaluation networks and education measurement efforts also helps to build and promote capacity in national institutions.

Promote the development of evaluation capacity at the local authority level

In highly decentralised education systems, evaluation capacity at the local authority level is crucial in promoting an effective evaluation culture. Where there is a wide variation in evaluation capacity among local authorities, establishing reporting requirements may be an effective stimulus to develop evaluation practices. For example, local authorities may be required to produce an annual evaluation report of schools in their system and to set objectives and development priorities for their school system. Formulating evaluation and assessment competency profiles for local officials carrying evaluation responsibilities is also a way to clarify expectations. A review of existing approaches to quality assurance by different local authorities would be a first step to identify examples of robust quality assurance practices. There is also the option of promoting and supporting local authority networks on different aspects of quality assurance and ensuring that results are fed back to build knowledge and capacity throughout the education system.

Ensure objectivity and credibility in education system evaluation activities

Education system evaluation plays an important role in providing a firm evidence base for policy making. A commitment to evidence-based policy making is commendable and can be demonstrated by ensuring objective and credible evaluations. This requires sufficient technical capacity to undertake education system evaluation and may involve establishing specific bodies competent in different aspects of education system evaluation, e.g. conducting research in priority areas, evaluating the implementation of major programmes or policies, developing national assessments, conducting external school evaluations, etc.

A way to raise credibility for evaluation activities is to give a clear mandate to different competent authorities or independent bodies to interpret evidence on the education system and to commission further evaluations or reviews accordingly. The findings of any commissioned reviews or evaluations should be published and feed into the wider public debate.

Reporting and use of results

In many education systems there are quite sophisticated measures of system performance in place which provide much information. This offers opportunities to engage stakeholders in supporting improvements across the school system. Typically, in OECD countries, while large amounts of data are collected from schools and comparable student results are available, there is room to strengthen the analysis and mobilisation of such information for system monitoring and improvement.

Strengthen analysis of education system evaluation results for planning and policy development

A priority should be the strengthening of the analysis for educational planning and policy development. It is clear that, in most countries, considerably more analysis and research could be conducted with the available data. Education authorities should promote analytical studies and innovative research about key issues such as the factors which explain student performance and the impact of the socio-economic background on student results. Education authorities could also sponsor research undertaken by independent researchers which is deemed useful for educational policy.

Another possibility is to require evaluation agencies or external school evaluation bodies to include thematic national evaluations among their responsibilities. These comprehensive reviews of issues such as science education, professional development of teachers, bullying and harassment in schools, and the teaching of mathematics, would involve reviewing practices across a sample of schools in the country and the production of a national level report. This would serve to inform the development and implementation of policy and practice.

Communicate key results of education system evaluation to stakeholders

Education authorities should promote measures to more effectively communicate results from the national monitoring system to encourage their use by different stakeholders. In many countries, there is room to improve the accessibility of system evaluation results. Often information on the education system is dispersed in components such as an education database, reporting on major education indicators, reporting on student outcomes and thematic reports on education. An option is to bring together the information into a single platform. Further, some systems publish a single report that compiles major system evaluation results to summarise the state of school education in the country. Another option is to establish a protocol to share data among key stakeholders in system evaluation – this may include data that are not available to the public, but that can be analysed and used, for example, for school or local government reviews. At the same time, it is important to include clear methodological documentation and full access to micro-level data for use by the research community and to include where necessary confidentiality requirements.

Also attention is required to provide clear and timely reporting of results to different audiences. Giving high-quality feedback on system results is one way to maximise the use of results by stakeholders throughout the system. For example, databases and technical materials are useful for researchers, but clear key messages on major results are helpful for local government and – where available – schools will benefit from comprehensive feedback on student performance on national or international student assessments (e.g. by area assessed, by individual question, by class, by student group). The release of new system-level information should be accompanied by clear key messages on the major results and may be accompanied by national conferences with key stakeholders to discuss what can be learned from the results and possible implications for policy and practice.

Support feedback for local monitoring

Education authorities should devise strategies to optimise the use of system-level data by key stakeholders at the local level such as local education authorities and also feedback to schools (see Chapter 6). It may be necessary to establish a protocol to share data – this may include data that are not available to the public, but that can be analysed and used, for example, for school or local authority reviews. An electronic platform, with a wide range of information on major indicators for individual schools and the access to information about all schools within a local authority can be an effective management and monitoring tool for local authorities. To encourage the use of such information systems for monitoring progress at the local level, such a system may include some benchmarks set nationally to serve as a springboard for regions and schools to set their own local objectives and targets. Reporting should have a strong focus placed on developing benchmarking analyses which are trusted and valued by school leadership. This means they must be based on reliable data but also that they should facilitate "fair and

meaningful" comparisons between schools, that is, by providing adequate contextual information on schools to be able to compare schools with similar contexts.

Local education authorities could possibly take a central feedback role. As they are closer to the local level than the national authority, local authorities could use school reporting data as a basis for engaging in meaningful discussions with schools and their leadership. To further strengthen the role of local education authorities, the national level authority could also consider disaggregating system level data to meaningful groupings to allow a more fine-grained analysis (e.g. by region, by school board, by school pedagogical network or community, etc.).

References

ACARA (Australian Curriculum, Assessment and Reporting Authority) 2011, *Measurement Framework for Schooling in Australia 2010*, Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA). *www.acara.edu.au/verve/_resources/Measurement_Framework_for_Schooling_in_Australia_2010.pdf*.

ADQS (Agency for the Development of Quality in Schools) (2011) *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Luxembourg,* Ministry of National Education and Vocational Training, Luxembourg, *www.oecd.org/edu/evaluationpolicy*.

Anderson, P. and G. Morgan (2008), *Developing Tests and Questionnaires for a National Assessment of Educational Achievement*, The World Bank, Washington, D.C.

Australian National Audit Office/CPA Australia (2008), *Monitoring and Reporting Financial and Non-financial Performance of Australian Government Organisations*, Issues Paper Series No. 5, February, *www.canberra.edu.au/corpgov-aps/pub/issuespaper5.pdf*.

Bialecki, I., S. Johnson, and G. Thorpe (2002), "Preparing for national monitoring in Poland", *Assessment in Education: Principles, Policy and Practice*, Vol. 9(2), pp. 221-236.

Breakspear, S. (2012), "The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance", *OECD Education Working Papers*, No. 71, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Brejc, M., M. Sardoč and D. Zupanc (2011), *Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Slovenia,* National School for Leadership in Education in cooperation with the National Examinations Centre, *www.oecd.org/edu/evaluationpolicy*.

Canadian Statistics Education Council (2012), *Education Indicators in Canada: An International Perspective 2012*, Statistics Canada, *www.statcan.gc.ca/pub/81-604-x/81-604-x2012001-eng.htm*.

Carvalho, L.M. and E. Costa (2009), "Production of OECD's programme for international student assessment (PISA), KNOWandPOL orientation 3 supra-national instruments WP11", *Knowledge and Policy in Education and Health Sectors*.

CITO (2009), *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2008*, Jaarlijks Peilingsonderzoek van het Onderwijsniveau, Technische rapportage, Cito B.V., Arnhem.

CITO (2008), *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON*, Stichting Cito Instituut voor Toetsontwikkeling, Arnhem.

CMEC (2010), *A Framework for Statistics on Learning and Education in Canada*, Council of Ministers of Education, Canada (CMEC).

CMEC (2008), *Pan-Canadian Assessment Program – PCAP-13 2007: Report on the Assessment of 13-Year-Olds in Reading, Mathematics, and Science*, Council of Ministers of Education, Canada.

Danish Ministry of Education and Rambøll (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Update to the Country Background Report for Denmark*, National Agency for Quality and Supervision, *www.oecd.org/edu/evaluationpolicy*.

DEEWR (Department of Education, Employment and Workplace Relations) (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Australia, www.oecd.org/edu/evaluationpolicy*.

Delvaux, B. and E. Mangez (2008), "Towards a sociology of the knowledge-policy relation – literature review, integrative report, September 2008", *Knowledge and Policy in Education and Health Sectors*.

Dutch Ministry of Education, Culture and Science (2012), *Key Figures 2007-2011 Education, Culture and Science*, Ministry of Education Culture and Science, The Hague.

Egelund, N. (2008), "The value of international comparative studies of achievement – a Danish perspective", *Assessment in Education: Principles, Policy & Practice*, Vol. 15(3), pp. 245-251.

Englert, K., D. Fries, M. Martin-Glenn and B. Douglas (2007), "Accountability systems: A comparative analysis of superintendent, principal and teacher perceptions", *International Journal of Education Policy and Leadership*, Vol. 2 (4).

Eurydice (2009), *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*, Education, Audiovisual and Culture Executive Agency, Brussels. *http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf*.

Fournier, G. and D. Mildon (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Canada*, Council of Ministers of Education, Canada (CMEC).

Goldstein, H. and S. Thomas (2008), "Reflections on the international comparative surveys debate", *Assessment in Education: Principles, Policy & Practice*, Vol. 15:3, pp. 215-222.

Greaney, V. and T. Kellaghan (2008), *Assessing National Achievement Levels in Education*, The World Bank, Washington, D.C.

Green, S. and T. Oates (2009), "Considering alternatives to national assessment arrangement in England: Possibilities and opportunities", *Educational Research*, Vol. 51, No. 2, June 2009, NFER, pp. 229-245.

Grek, S. (2009), "Governing by numbers: The PISA effect in Europe", *Journal of Education Policy*, Vol. 24(1), pp. 23-37.

Gür, B.S., Z. Çelik and M. Özoglu (2011), "Policy options for Turkey: A critique of the interpretation and utilisation of PISA results in Turkey", *Journal of Education Policy*, Vol. 27(1), pp. 1-21.

Hajdúková, V., M. Šnídlová, A. Vykydalová, T. Lukács, A. Mentel, M. Pokorný, Z. Juščáková and R. Kanovská (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Slovak Republic,* National Institute for Certified Educational Measurements (NÚCEM), Bratislava.

Hamilton, L. and D. Koretz (2002), "Tests and their use in test-based accountability systems", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education,* RAND Publishing, Santa Monica, California.

Harlen, W. (2007), "Criteria for evaluating systems for student assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 15-28.

Hopkins, D., D. Pennock, J. Ritzen, E. Ahtaridou and K. Zimmer (2008), "External evaluation of the policy impact of PISA", OECD document EDU/PISA/GB(2008)35/REV1, OECD, Paris.

House of Commons Children, Schools and Families Committee (2008), *Testing and Assessment – Third Report of Session 2007–08 Volume I: Report*, together with formal minutes, The Stationery Office, London, *www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/169/169.pdf*.

INEE (2006b), "La difusión de los resultados de las evaluaciones", *Los Temas de la Evaluación No. 20*, Colección de Folletos, Instituto Nacional para la Evaluación de la Educación, Mexico, D.F.

Jensen, T.P. and D. Andersen (2006), "Participants in 2000 – four years later", in J. Medjing and A. Roe (eds.), *Northern Lights on PISA 2003: A Reflection from the Nordic Countries*, Nordic Council of Ministers, Copenhagen.

Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim and JunghaPark (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea,* Korean Educational Development Institute (KEDI), Seoul, *www.oecd.org/edu/evaluationpolicy*.

KL (Local Government Denmark) (2009), "Partnerskab om Folkeskolen – Kort og godt", Kommuneforlaget A/S, *www.kl.dk/ImageVault/Images/id_41939/ImageVaultHandler.aspx*.

Koretz, D. (2010), *Implications of Current Policy for Educational Measurement*, Education Testing Service (ETS).

Le, V. and S. Klein (2002), "Technical criteria for evaluating tests", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education*, RAND Publishing, Santa Monica, California.

Levin, J. A. and A. Datnow (2012), "The principal role in data-driven decision making: Using case-study data to develop multi-mediator models of educational reform", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 23:2, 179-201, Routledge.

Masters, G.N. (2012), *Measuring and Rewarding School Improvement*, a discussion paper commissioned by the Commonwealth Department of Education, Employment and Workplace Relations (DEEWR).

MCEETYA (2008), *Measurement Framework for National Key Performance Measures*, Performance Measurement and Reporting Taskforce, Ministerial Council on Education, Employment, Training and Youth Affairs.

MCEETYA (2000), *National Report on Schooling in Australia 2000*, Ministerial Council on Education, Employment, Training and Youth Affairs, *http://cms.curriculum.edu.au/anr2000/index.htm*.

McGaw, B. (2008), "The role of the OECD in international comparative studies of achievement", *Assessment in Education: Principles, Policy & Practice*, Vol. 15, No. 3, pp. 223-243.

Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium,* Brussels, *www.oecd.org/edu/evaluationpolicy*.

Ministry of Education (2009), *Statement of Intent 2009-2014*, Ministry of Education, New Zealand, *www.minedu.govt.nz/~/media/MinEdu/Files/TheMinistry/StatementOfIntent2009.pdf*.

Morris, A. (2011), "Student standardised testing: Current practices in OECD countries and a literature review", *OECD Education Working Papers,* No. 65, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Nusche, D., D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

OECD (2012), *Quality Framework and Guidelines for OECD Statistical Activities,* Version 2011/1, Statistics Directorate, OECD, Paris, *http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=std/qfs(2011)1&doclanguage=en*.

OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, Paris. *http://dx.doi.org/10.1787/eag-2011-en*.

OECD (2010a), *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume 1)*, OECD Publishing, Paris.

OECD (2010b), *Educational Information Systems for Innovation and Improvement: Workshop Summary Report*, 14-15 October 2010, Directorate for Education, Centre for Educational Research and Innovation (CERI) Governing Board, OECD, Paris.

OECD (2010c), *Pathways to Success – How Knowledge and Skills at Age 15 Shape Future Lives in Canada*, OECD, Paris, *www.pisa.oecd.org/dataoecd/59/35/44574748.pdf*.

OECD (2007a), *OECD Budget Practices and Procedures Database Phase II – Final Glossary*, 19 December 2006, OECD, Paris, *www.oecd.org/dataoecd/30/46/39466131.pdf*.

OECD (2007b), *Evidence in Education: Linking Research and Policy*, OECD Publishing, Paris.

OECD (2004), *Review of National Policies for Education – Denmark: Lessons from PISA 2000*, OECD, Paris.

OECD (2001), *Knowledge and Skills for Life – First Results from PISA 2000*, OECD, Paris.

Ofqual (2008), *The Regulation of Examinations and Qualifications: An International Study*, Office of the Qualifications and Examinations Regulator, England, *http://inca.org.uk/ofqual-08-3736_regulation_of_examinations_and_qualifications.pdf*.

Perie, M. and J. Park (2007), *Key Elements for Educational Accountability Models* – A paper commissioned by the Council of Chief State School Officers Accountability Systems and Reporting State Collaborative, The Council of Chief State School Officers, Washington, D.C.

Portuguese Ministry of Education and Science (forthcoming), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Portugal*, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Rautalin, M. and P. Alasuutari (2007), "The curse of success: The impact of OECD's Programme for International Student Assessment on the discourses of the teaching profession in Finland", *European Educational Research Journal*, Vol. 6, No. 4, pp. 348-63.

Regeringen (2010), *Danmark 2020 Viden > vækst > velstand > velfærd*, Statsministeriet, Copenhagen.

Rosenkvist, M. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers,* No. 54, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012a), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012b), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Santiago, P., I. McGregor, D. Nusche, P. Ravela and D. Toledo (2012c), *OECD Reviews of Evaluation and Assessment in Education: Mexico*, OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Sauvageot, C. (2008), "International comparisons – Introduction", *Éducation et formations, no. 78* November 2008, Ministry of National Education, Ministry of Higher Education and Research, Direction de l'évaluation, de la prospective et de la performance (DEPP), Imprimerie Moderne de l'Est, Paris.

Scheerens, J., M. Ehren, P. Sleegers and R. de Leeuw (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Netherlands,* University of Twente, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., M. Ehren, P. Santiago and C. Tamassia (2012), *OECD Reviews of Evaluation and Assessment in Education: Luxembourg,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy.*

Shewbridge, C., M. Hulshof, D. Nusche and L. Stenius Stæhr (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Northern Ireland, United Kingdom*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

Shewbridge, C., D. Nusche, P. Wright, P. and J. van Bruggen (forthcoming), *OECD Reviews of Evaluation and Assessment in Education: Slovak Republic*, OECD Publishing, Paris, forthcoming at *www.oecd.org/edu/evaluationpolicy*.

SOU (2009), "Att nå ut och nå ända fram, Hur tillgången till policyinriktad utvärdering och forskningsresultat inom utbildningsområdet kan tillgodoses", Statens Offentliga Utrewdningar, Swedish Government Official Reports, SOU 2009:94.

Specht, W. and F. Sobanski (2012), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Austria*, Bundesministerium für Unterricht, Kunst und Kultur (BMUKK,), Vienna, *www.oecd.org/edu/evaluationpolicy*.

Steering Committee for the Review of Government Service Provision (2010), *Review of the Report on Government Services' Performance Indicator Framework – Report to the Steering Committee for the Review of Government Service Provision*, prepared by the Independent Reference Group, August 2010, *www.pc.gov.au/gsp/independent-reference-group-report*.

Tamassia, C. and R. Adams (2009), "International assessments and indicators – How will assessments and performance indicators improve educational policies and practices in a globalised society?", in K. Ryan (2009), *The SAGE International Handbook of Educational Evaluation*, SAGE Publications, Thousand Oaks, California.

U.S. Department of Education (2010), *Use of Education Data at the Local Level – From Accountability to Instructional Improvement*, U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Washington D.C., *www2.ed.gov/rschstat/eval/tech/use-of-education-data/use-of-education-data.pdf*.

Visscher, A.J. and R. Coe (2003), "School performance feedback systems: Conceptualisation, analysis and reflection", *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, Vol. 14:3, pp. 321-349, Routledge.

Wayman, J.C. (2005), "Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection", *Journal of Education for Students Placed at Risk,* Vol. 10, No. 3, 295–308.

Zucker, S. (2003), *Fundamentals of Standardized Testing*, Pearson Education, *www.pearsonassessments.com*, accessed 12 July 2011.

Annex A. How the Review on Evaluation and Assessment Frameworks for Improving School Outcomes was conducted

Background to the Review

Quality in education is increasingly judged on the basis of effective learning outcomes. Countries use a range of techniques for the evaluation and assessment of students, teachers, school leaders, schools and education systems with the objective of assessing whether the school system is delivering good performance and providing feedback for improvement in student outcomes. Many countries test samples and/or all students at key points and also formally review the work of individual teachers, school leaders and schools. However, countries often face challenges in bringing the different elements of evaluation and assessment together into a coherent and comprehensive strategy, within which each element is fit for purpose and contributes effectively to improving learning outcomes.

Against this background, the OECD Education Policy Committee launched the Review on Evaluation and Assessment Frameworks for Improving School Outcomes in late 2009 to provide analysis and policy advice to countries on how assessment and evaluation can be embedded within a consistent framework to bring about real gains in performance across the school system.

Purpose of the Review

The Review was designed to respond to the strong interest in evaluation and assessment issues evident at national and international levels. The Review looked at the various components of assessment and evaluation frameworks that countries use with the objective of improving student outcomes. These include student assessment, teacher appraisal, school evaluation, school leader appraisal, system evaluation and also other elements such as the evaluation of an educational programme.

The overall purpose of the Review was to explore how evaluation and assessment frameworks can be used to improve the quality, equity and efficiency of school education. The overarching policy question was "How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools?" The Review further concentrated on five key issues for analysis: *(i)* designing a systemic framework for evaluation and assessment; *(ii)* ensuring the effectiveness of evaluation and assessment procedures; *(iii)* developing competencies for evaluation and for using feedback; *(iv)* making the best use of evaluation results; and *(v)* implementing evaluation and assessment policies. The analysis focused on primary and secondary levels of education.[1]

The main objectives were to:

- synthesise research-based evidence on the impact of evaluation and assessment strategies and disseminate this knowledge among countries
- identify innovative and successful policy initiatives and practices
- facilitate exchanges of lessons and experiences among countries
- identify policy options for policy makers to consider.

The Review's purposes, analytical framework and methodology are detailed in the OECD (2009) document entitled "Design and Implementation Plan for the Review" [EDU/EDPC(2009)3/REV1], available from *www.oecd.org/edu/evaluationpolicy*.

The Review was intended to extend and add value to the existing body of international work on educational evaluation. The importance of evaluation and assessment policies is reflected in a wide variety of other OECD activities, with which the Review collaborated. These include work by the Public Governance and Territorial Development Directorate on Public Governance Reviews and Public Sector Evaluation, the Programme for International Student Assessment (PISA), the Teaching and Learning International Survey (TALIS), CERI's Innovative Learning Environments (ILE) project, the INES Network for the Collection and the Adjudication of System-Level Descriptive Information on Educational Structures, Policies and Practices (NESLI), CERI's work on Longitudinal Information Systems as part of its Innovation Strategy for Education and Training, CERI's work on Governing Complex Education Systems and the Centre for Effective Learning Environments' (CELE) work on evaluating quality in educational facilities.

The growing attention paid to evaluation and assessment issues is also evident in the work of other international organisations. The OECD Review was therefore conducted in co-operation with a range of international organisations to reduce duplication and develop synergies, in particular: the European Commission, Eurydice, the Standing International Conference of Inspectorates (SICI), UNESCO and the World Bank. Social partners were also involved through the contribution of the Trade Union Advisory Committee to the OECD (TUAC) and the Business and Industry Advisory Committee to the OECD (BIAC).

Methodology and country participation

The Review methodology was based on countries working collaboratively with the Secretariat. It involved examining country-specific issues and policy approaches in using assessment and evaluation to improve student outcomes, and placing these experiences within a broader analytical framework to generate insights and findings relevant to OECD countries as a whole. The collaborative approach provided countries with an opportunity to learn more about themselves by examining their experiences against those of other countries. It was also intended to add to the broader knowledge base by accumulating international evidence on the impact of policy reforms, and the circumstances under which they work best.

The work was undertaken through a combination of desk-based analysis and country reviews. This was complemented with the meetings of the OECD Group of National Experts on Evaluation and Assessment (see below) to discuss progress and share experiences. The work was organised in three phases:

- An analytical phase, to draw together evidence-based policy lessons from international data, research and analysis. The analytical phase used several means – country background reports, literature reviews and data analyses – to analyse the factors that shape evaluation and assessment in school systems and possible policy responses.
- A country review phase, to provide policy advice to individual countries tailored to the issues of interest in those countries, on the basis of the international evidence base, combined with evidence obtained by a team of experts visiting the country.
- A synthesis phase, with the preparation of a final synthesis report to blend analytic and review evidence and provide overall policy conclusions.

Participating countries

The following countries were actively engaged in the Review:

Twenty-six systems (25 countries) prepared a Country Background Report: Australia, Austria, Belgium (Flemish Community), Belgium (French Community), Canada,[2] Chile, the Czech Republic, Denmark, Finland, France, Hungary, Iceland, Ireland, Italy, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Sweden and the United Kingdom (Northern Ireland).

Fifteen countries also opted for a Country Review: Australia, Belgium (Flemish Community), Chile, the Czech Republic, Denmark, Italy, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Portugal, the Slovak Republic, Sweden and the United Kingdom (Northern Ireland).

In addition to the countries mentioned above, the following countries provided detailed information about the features of their evaluation and assessment framework (synthesis tables, see below): Estonia, Israel and Spain.

Governance of the Review

The project was overseen by the Group of National Experts (GNE) on Evaluation and Assessment, which was established as a subsidiary body of the Education Policy Committee in order to guide the methods, timing and principles of the Review and to allow countries to share information and experience on evaluation and assessment as well as keeping in touch with emerging findings of the exercise. The GNE on Evaluation and Assessment was chaired by Mr. Gábor Halász, Professor, Eötvös Loránd University, Hungary and had as Vice-Chairs Mr. Éamonn Murtagh, Assistant Chief Inspector, Department of Education and Skills, Ireland, and Mr. Kwangho Kim, Research Fellow, Korean Educational Development Institute, Korea.

During the Review, the GNE on Evaluation and Assessment held four official meetings at the OECD Conference Centre in Paris. These were open to all OECD member countries and observers to the Education Policy Committee as well as to the Trade Union Advisory Committee to the OECD (TUAC) and the Business and Industry Advisory Committee to the OECD (BIAC). The objectives and main issues treated at each meeting are described in Table A.1 below.

Table A.1 Meetings of the Group of National Experts on Evaluation and Assessment

Date	Main issues treated
1-2 October 2009	Discussion of key policy issues. Countries agreed on the framework, scope and process for the Review and identified the major issues for investigation. The Review was officially launched.
9-10 September 2010	Countries exchanged ideas about key issues in evaluation and assessment policy on the basis of analytical work developed by the Review, shared country experiences on specific issues and discussed the progress of the Review.
13-14 October 2011	Countries reviewed the extended outline for the preparation of the final synthesis report from the Review and discussed the progress of the Review activities.
31 May – 1 June 2012	Following an update on developments in the countries and progress made within the Review, countries discussed the structure and the first draft of the Synthesis Report from the Review and shared country experiences on key themes of the report.

National co-ordinators

Each participating country appointed a national co-ordinator, who was responsible for: communications with the OECD Secretariat and within the country about the Review; ensuring that the Country Background Report was completed on schedule; liaising with the OECD Secretariat about the organisation of the review team visit, for those countries which opted for a country review; attending meetings of the Group of National Experts on Evaluation and Assessment; co-ordinating country feedback on draft materials; and assisting with dissemination activities. National co-ordinators are listed in Table A.2.

Table A.2 National co-ordinators in participating countries

Country	National co-ordinator
Australia	Ms. Kristie van Omme, Department of Education, Employment and Workplace Relations Ms. Nina Downes, Department of Education, Employment and Workplace Relations
Austria	Mr. Florian Sobanski, Federal Ministry of Education, Arts and Culture
Belgium (Flemish Community)	Mrs. Miekatrien Sterck, Flemish Ministry of Education and Training Mrs. Micheline Scheys, Flemish Ministry of Education and Training
Belgium (French Community)	Mr. Etienne Gilliard, Ministry of Education of the French Community of Belgium
Canada	Ms. Antonella Manca-Mangoff, Council of Ministers of Education
Chile	Mr. Francisco Lagos M., Ministry of Education
Czech Republic	Ms. Jana Straková, Institute for Information on Education (until December 2011) Ms. Michaela Sojdrova, Czech Schools Inspectorate
Denmark	Mrs. Charlotte Rotbøll, Ministry of Children and Education
Finland	Ms. Kirsi Kangaspunta, Ministry of Education Mr. Tommi Karjalainen, Ministry of Education
France	Mrs. Florence Lefresne, Ministry of National Education Mr Bruno Trosseille, Ministry of National Education
Hungary	Mr. László Limbacher, Ministry of Human Relations (Ministry of Education and Culture until 2010) Professor Gábor Halász, Eötvös Lóránd University/ELTE
Iceland	Mr. Stefán Baldursson, Ministry of Education, Science and Culture
Ireland	Mr. Éamonn Murtagh, Department of Education and Skills
Italy	Mr. Paolo Sestito, National Institute for the Evaluation of the Education System (INVALSI) Mr. Daniele Vidoni, National Institute for the Evaluation of the Education System (INVALSI)

Table A.2 National co-ordinators in participating countries (*continued*)

Country	National co-ordinator
Korea	Mr. Kwangho Kim, Korea Educational Development Institute
Luxembourg	Ms. Amina Kafai, Ministry of National Education and Vocational Training
Mexico	Mr. Alejandro S. Ramírez Torres, Secretariat of Public Education (until December 2011) Ms. Florencia Martínez Becerra, Secretariat of Public Education
Netherlands	Mrs. Marian Hulshof, Inspectorate of Education
New Zealand	Ms. Ro Parsons, Ministry of Education
Norway	Ms. Vivi Bjelke, Directorate for Education and Training
Poland	Mr. Michal Sitek, Educational Research Institute
Portugal	Prof. Alexandre Ventura, Scientific Council for Teacher Evaluation (until June 2010) Prof. Natércio Afonso, University of Lisbon (until June 2011) Prof. Luísa Canto e Castro Loura, Ministry of Education and Science
Slovak Republic	Mr. Andrej Mentel, National Institute for Certified Educational Measurements (until July 2012) Mr. Martin Pokorny, National Institute for Certified Educational Measurements
Slovenia	Mrs. Mateja Brejc, National School for Leadership
Sweden	Ms. Kerstin Hultgren, Ministry of Education and Research Ms. Ann-Kristin Boström, Swedish National Agency for Education
United Kingdom (Northern Ireland)	Ms. Karen McCullough Department of Education for Northern Ireland Mr. Dale Heaney Department of Education for Northern Ireland

Country Background Reports

Information on countries' policies and practices was gathered through Country Background Reports (CBRs). The CBRs were prepared in response to a common set of issues and questions, and used a common framework to facilitate comparative analysis and maximise the opportunities for countries to learn from each other. The CBRs were a major source of material for this final report. The Guidelines for the preparation of CBRs are set out in a dedicated document [EDU/EDPC/EA(2009)1/REV1], also available on the Review website (*www.oecd.org/edu/evaluationpolicy*).

The CBRs were structured around the following main chapters:

1. The school system
2. The framework for evaluation and assessment
3. System evaluation
4. School evaluation
5. Teacher appraisal
6. Student assessment
7. Other types of evaluation and assessment

The work on the CBRs took place mainly between the end of 2009 and the end of 2011. Countries differed somewhat in the time they joined the Review and time needed to complete and publish their CBR. Countries also differed in the extent to which they were able to include current data and policy developments in their reports. Therefore the CBRs do not all refer to the same period, although most encompass developments up to about 2010. In late 2012 some countries prepared updates on their CBR for publication on the Review website.

The CBRs were intended for four main audiences: The Secretariat and OECD member and observer countries as an aid to sharing experiences, and identifying common problems and policy options; the team of external reviewers who visited the countries which opted for a Country Review; those interested in evaluation and assessment issues within the country concerned; and those interested in evaluation and assessment issues at international level and in other countries. All CBRs are available on the Review website: *www.oecd.org/edu/evaluationpolicy.*

Synthesis tables

In addition to the Country Background Reports, all countries provided detailed qualitative information on features of their evaluation and assessment framework through a questionnaire prepared by the OECD Secretariat. The data covered standardised national examinations, internal summative assessment, reporting of summative results, student formative assessment, teacher appraisal, employment status and career development of teachers, school leadership appraisal, employment status and career development of school leaders, education system evaluation and standardised national assessments. This information complemented other relevant information available through OECD's *Education at a Glance*. The information provided by countries was published in a set of tables included in this report.

Country Review Reports

Another major source of material for this report was the set of Country Review Reports prepared by the external review teams that visited countries taking part in the Country Review Strand. By providing an external perspective on evaluation and assessment policy issues in the countries concerned, the Country Review Reports were also intended to contribute to national discussions, as well as inform other countries about policy innovations underway. The Country Review Reports were also published as a publication series: *OECD Reviews of Evaluation and Assessment in Education*, in order to enhance the visibility of these country-specific outputs as part of the Review.

For each country visited, a team of up to five reviewers (including at least two OECD Secretariat members) analysed the Country Background Report and associated materials and subsequently undertook an intensive case study visit of about 8 days in length. The reviewers were selected in consultation with the country authorities to ensure that they had experience relevant to the main policy issues in the country concerned. The study visit aimed to provide the review team with a variety of perspectives on evaluation and assessment policy and included meetings with education authorities at national and sub-national levels; relevant agencies that deal with evaluation and assessment issues (e.g. national assessment agencies, school inspectorates, etc.); teacher professional organisations; teacher unions; parents' organisations; representatives of schools; representatives of school leaders; students' organisations; teacher educators; researchers with an interest in evaluation and assessment issues, and groups of students, teachers and school leaders at the schools visited. The objective was to accumulate sufficient information and understanding on which to base the analysis and policy recommendations.

The 15 review visits involved 30 external reviewers with a range of research and policy backgrounds. Overall, the external review teams visited about 90 schools and met with about 2 800 individuals to base their findings. Details on the country review visits are given in Table A.3. The Country Review Reports are published on the project website: *www.oecd.org/edu/evaluationpolicy*.

Table A.3 Country reviews and team members

Country	Review team
Sweden 4-11 May 2010	Deborah Nusche, OECD Secretariat (co-ordinator) Gábor Halász, University Eötvös Loránd, Hungary Janet Looney, formerly with the OECD; United States Paulo Santiago, OECD Secretariat Claire Shewbridge, OECD Secretariat
Luxembourg 31 May – 4 June 2010	Claire Shewbridge, OECD Secretariat (co-ordinator) Melanie Ehren, University of Twente; Netherlands Morten Rosenkvist, OECD Secretariat Paulo Santiago, OECD Secretariat Cláudia Tamassia, Educational Testing Service (ETS); Brazil
Australia 21-30 June 2010	Paulo Santiago, OECD Secretariat (co-ordinator) Graham Donaldson, former HM Senior Chief Inspector of Education in Scotland, United Kingdom Joan Herman, University of California – Los Angeles, United States Claire Shewbridge, OECD Secretariat
New Zealand 23-30 August 2010	Deborah Nusche, OECD Secretariat (co-ordinator) Dany Laveault, University of Ottawa, Canada John MacBeath, University of Cambridge, United Kingdom Paulo Santiago, OECD Secretariat
Denmark 5-12 October 2010	Claire Shewbridge, OECD Secretariat (co-ordinator) Eunice Jang, University of Toronto, Canada Peter Matthews, University of London, United Kingdom Paulo Santiago, OECD Secretariat
Norway 8-15 December 2010	Deborah Nusche, OECD Secretariat (co-ordinator) Lorna Earl, formerly with the University of Toronto, Canada William Maxwell, Her Majesty's Senior Chief Inspector of Education in Scotland, United Kingdom Claire Shewbridge, OECD Secretariat
Belgium (Flemish Community) 25 January – 1 February 2011	Claire Shewbridge, OECD Secretariat (co-ordinator) Marian Hulshof, Inspectorate of Education, Netherlands Deborah Nusche, OECD Secretariat Louise Stoll, University of London, United Kingdom
Portugal 23 February – 2 March 2011	Paulo Santiago, OECD Secretariat (co-ordinator) Graham Donaldson, former HM Senior Chief Inspector of Education in Scotland, United Kingdom Anne Looney, National Council for Curriculum and Assessment, Ireland Deborah Nusche, OECD Secretariat
Czech Republic 29 March – 5 April 2011	Paulo Santiago, OECD Secretariat (co-ordinator) Alison Gilmore, University of Otago, New Zealand Deborah Nusche, OECD Secretariat Pamela Sammons, University of Oxford, United Kingdom
Chile 2-9 November 2011	Paulo Santiago, OECD Secretariat (co-ordinator) Francisco Benavides, OECD Secretariat Charlotte Danielson, Educational Consultant, United States Laura Goe, Educational Testing Service, United States Deborah Nusche, OECD Secretariat
Mexico 7-15 February 2012	Paulo Santiago, OECD Secretariat (co-ordinator) Isobel McGregor, formerly with Her Majesty's Inspectorate of Education in Scotland, United Kingdom Deborah Nusche, OECD Secretariat Pedro Ravela, Catholic University of Uruguay, Uruguay Diana Toledo, OECD Secretariat

Table A.3 Country reviews and team members (*continued*)

Country	Review team
Slovak Republic 27 March – 3 April 2012	Claire Shewbridge, OECD Secretariat (co-ordinator) Johan van Bruggen, formerly with the Dutch Inspectorate of Education, Netherlands Deborah Nusche, OECD Secretariat Paul Wright, formerly with the Qualifications and Curriculum Development Agency, United Kingdom
United Kingdom (Northern Ireland) 26 February – 5 March 2013	Claire Shewbridge, OECD Secretariat (co-ordinator) Marian Hulshof, Inspectorate of Education, Netherlands Deborah Nusche, OECD Secretariat Lars Stenius Staehr, Testing Consultant for the Ministry of Education, Denmark
Netherlands 4-11 June 2013	Deborah Nusche, OECD Secretariat (co-ordinator) Paulo Santiago, OECD Secretariat Two external reviewers (not defined at the time of printing this report)
Italy (dates not defined at the time of printing this report)	Paulo Santiago, OECD Secretariat (co-ordinator) Claire Shewbridge, OECD Secretariat Two external reviewers (not defined at the time of printing this report)

Commissioned and background papers

The Review was also enriched through the following background papers and literature reviews which address particular issues in depth:

Literature Reviews

- *Making Student Assessment Inclusive*, by Anne Nayral de Puybusque and edited by Thomas Radinger, forthcoming
- *School Leadership Appraisal: Current Practices in OECD Countries and a Literature Review*, by Thomas Radinger, forthcoming
- *Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence*, by Stefanie Dufaux, 2012
- *Student Standardised Testing: Current Practices in OECD Countries and a Literature Review*, by Allison Morris, 2011
- *Using Student Test Results for Accountability and Improvement: A Literature Review*, by Morten Rosenkvist, 2010
- *School Evaluation: Current Practices in OECD Countries and a Literature Review*, by Violaine Faubert, 2009
- *Teacher Evaluation: Current Practices in OECD Countries and a Literature Review*, by Marlène Isoré, 2009

Commissioned and background papers

- *Alignment in Complex Education Systems: Achieving Balance and Coherence*, by Janet Looney, 2011
- *Integrating Formative and Summative Assessment: Progress toward a Seamless System?*, by Janet Looney, 2011

- *Evaluation and Assessment Frameworks for Improving School Outcomes: Common Policy Challenges*, 2011
- *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*, by Paulo Santiago and Francisco Benavides, 2009

Dissemination

The Review had a strong emphasis on dissemination from the outset. To facilitate dissemination and encourage feedback, all project documents and outputs were published on the Review's website *www.oecd.org/edu/evaluationpolicy*. Throughout the Review, the OECD Secretariat made a large number of presentations about the project to a wide range of meetings and groups of visitors to the OECD.

An international conference was hosted by the Ministry of Education and Research of Norway and the Norwegian Directorate for Education and Training in Oslo on 11-12 April 2013 to conclude the Review and launch this final report. A significant number of countries also organised national events to discuss both the international results from the Review and the conclusions of specific country reviews.

Notes

1. The scope of the Review did not include early childhood education and care, apprenticeships within vocational education and training, and adult education.
2. The Country Background Report for Canada focuses on policies and practices of the Atlantic provinces: Newfoundland and Labrador, Nova Scotia, Prince Edward Island, and New Brunswick.

Annex B. Summary of policy directions from the Review on Evaluation and Assessment Frameworks for Improving School Outcomes

Table B.1 Policy directions for the evaluation and assessment framework

Domain	Policy directions
Governance	• Integrate the evaluation and assessment framework • Align the evaluation and assessment framework with educational goals and student learning objectives • Secure links to the classroom and draw on teacher professionalism • Give a prominent role to independent evaluation agencies • Promote national consistency while giving room for local diversity • Integrate the non-public sector in the overall evaluation and assessment framework
Design and procedures	• Ensure core components are sufficiently developed within the evaluation and assessment framework • Establish articulations between components of the evaluation and assessment framework • Place the students at the centre of the evaluation and assessment framework • Build on some key principles to effectively implement evaluation and assessment
Capacity	• Sustain efforts to improve capacity for evaluation and assessment • Improve the articulation between levels of authority and assure support from the centre
Use of results	• Maintain sound knowledge management within the overall evaluation and assessment framework • Commit to the use of evidence for policy development
Implementation	• Anticipate potential implementation difficulties • Engage stakeholders and practitioners in the design and implementation of evaluation and assessment policies • Communicate the rationale for reform • Use pilots before full implementation and review implementation • Ensure adequate capacity and sufficient resources

Table B.2 Policy directions for student assessment

Domain	Policy directions
Governance	• Establish a coherent framework for student assessment • Develop clear goals and learning progressions to guide student assessment • Ensure a good balance between formative and summative assessment • Establish safeguards against an overreliance on standardised assessments • Share responsibilities for the governance and implementation of assessment frameworks
Procedures	• Draw on a variety of assessment types to obtain a rounded picture of student learning • Support effective formative assessment processes • Clarify and illustrate criteria to judge performance in relation to national goals • Ensure the consistency of assessment and marking across schools • Promote assessment formats that capture valued key competencies • Build on innovative approaches developed in particular education sectors • Tap into the potential of ICT to develop sophisticated assessment instruments • Ensure that student assessment is inclusive and responsive to different learner needs

Table B.2 Policy directions for student assessment (*continued*)

Domain	Policy directions
Capacity	• Put the learner at the centre and build students' capacity to engage in their own assessment • Maintain the centrality of teacher-based assessment and promote teacher professionalism • Identify assessment priority areas for teacher initial education and professional development • Use teacher appraisal and school evaluation processes to help teachers develop their assessment capacity
Reporting and use of results	• Develop clear reporting guidelines • Engage parents in education through adequate reporting and communication • Ensure transparency and fairness when using assessment results for high-stakes decisions • Promote the regular use of assessment results for improvement

Table B.3 Policy directions for teacher appraisal

Domain	Policy directions
Governance	• Clarify the purposes of teacher appraisal and ensure that it fits national education objectives • Resolve tensions between the developmental and accountability functions of teacher appraisal • Establish a coherent framework for teacher appraisal • Establish a mandatory probationary period for new teachers • Consolidate regular developmental appraisal at the school level • Establish periodic career-progression appraisal involving external evaluators • Create a teacher career structure with distinct pathways • Address the challenges of implementation
Procedures	• Establish teaching standards to guide teacher appraisal and professional development • Use multiple instruments and sources of evidence • Provide support for effective classroom observations • Establish safeguards against simplistic use of student results for teacher appraisal • Ensure that student feedback to teachers is used for formative purposes
Capacity	• Prepare teachers for their role in appraisal processes • Strengthen the capacity of school leaders for teacher appraisal • Ensure that designated evaluators are qualified for their role • Build central expertise to continuously improve teacher appraisal policies and practices
Use of results	• Ensure that teacher appraisal feeds into professional development and school development • Establish feedback loops between teacher appraisal systems and initial teacher education • Establish links between teacher appraisal and career advancement decisions • Consider the use of non-monetary rewards as a complementary tool to recognise teachers • Ensure that underperformance is identified and adequately addressed

Table B.4 Policy directions for school evaluation

Domain	Policy directions
Governance	• Clarify the role and purpose of school evaluation within the wider evaluation and assessment framework • Ensure the focus for school evaluation is the improvement of teaching, learning and student outcomes • Evaluate and adapt external school evaluation to reflect the maturity of the school evaluation culture • Raise the profile of school self-evaluation • Align external school evaluation with school self-evaluation
Procedures	• Develop nationally agreed criteria for school quality to guide school evaluation • Develop appropriate resources for school self-evaluation • Ensure a strong evidence base for external school evaluation and appropriate analysis tools • Ensure transparency in external school evaluation procedures
Capacity	• Ensure the credibility of external evaluators and enhance their objectivity and coherence • Ensure sufficient capacity and retraining as necessary to fit the approach to external school evaluation • Strengthen school principals' capacity to stimulate an effective school self-evaluation culture • Promote the engagement of all school staff and students in school self-evaluation • Promote peer learning among schools
Reporting and use of results	• Optimise the feedback of nationally collected data to schools for self-evaluation and development planning • Promote the wider use of the results of external school evaluation • Ensure the systematic follow-up of external school evaluations • Report a broad set of school performance measures with adequate contextual information

Table B.5 Policy directions for the appraisal of school leaders

Domain	Policy directions
Governance	• Promote the effective appraisal of school leaders within the broader assessment and evaluation framework while giving room for local diversity • Clarify the purposes of school leader appraisal
Procedures	• Develop a common leadership framework or set of professional standards for school leaders • Promote the appraisal of pedagogical leadership together with scope for local adaptation • Promote the appraisal of school leaders' competencies for monitoring, evaluation and assessment • Consider school leaders' efforts to distribute leadership and enhance teacher leadership within schools and assume leadership responsibilities beyond their school borders as an integral part of appraisal • Promote the use of multiple instruments and sources of evidence
Capacity	• Build capacity for effective school leader appraisal • Promote school leader appraisal as an opportunity for peer learning
Use of results	• Ensure school leader appraisal informs professional development • Consider the development of a career structure and career advancement opportunities to reward successful school leaders

Table B.6 Policy directions for education system evaluation

Domain	Policy directions
Governance	• Ensure a broad concept of education system evaluation within the evaluation and assessment framework • Ensure policy making is informed by high-quality measures, but not driven by their availability • Situate education system evaluation in the broader context of public sector performance requirements
Procedures	• Develop a national education indicator framework • Design a national strategy to monitor student learning standards • Ensure the collection of qualitative information on the education system • Assure the monitoring of changes over time and progress of particular student cohorts • Ensure collection of adequate contextual information to effectively monitor equity
Capacity	• Establish and secure capacity for education system evaluation • Promote the development of evaluation capacity at the local authority level • Ensure objectivity and credibility in education system evaluation activities
Reporting and use of results	• Strengthen analysis of education system evaluation results for planning and policy development • Communicate key results of education system evaluation to stakeholders • Support feedback for local monitoring

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(91 2013 02 1 P) ISBN 978-92-64-19064-1 – No. 60541 2013

OECD Reviews of Evaluation and Assessment in Education

Synergies for Better Learning

AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT

How can assessment and evaluation policies work together more effectively to improve student outcomes in primary and secondary schools? Countries increasingly use a range of techniques for student assessment, teacher appraisal, school evaluation, school leader appraisal and education system evaluation. However, they often face difficulties in implementing evaluation and assessment policies. This may arise as a result of poor policy design, lack of analysis of unintended consequences, little capacity for school agents to put evaluation procedures into practice, lack of an evaluation culture, or deficient use of evaluation results.

This report provides an international comparative analysis and policy advice to countries on how evaluation and assessment arrangements can be embedded within a consistent framework to improve the quality, equity and efficiency of school education. It builds upon a major 3-year review of evaluation and assessment policies in 28 countries, the *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*. As well as analysing strengths and weaknesses of different approaches, the report provides recommendations for improvement including how results should be incorporated into policy and practice.

Contents

Chapter 1. The focus on evaluation and assessment
Chapter 2. Trends in evaluation and assessment
Chapter 3. The evaluation and assessment framework: Embracing a holistic approach
Chapter 4. Student assessment: Putting the learner at the centre
Chapter 5. Teacher appraisal: Enhancing teacher professionalism
Chapter 6. School evaluation: From compliancy to quality
Chapter 7. The appraisal of school leaders: Fostering pedagogical leadership in schools
Chapter 8. Education system evaluation: Informing policies for system improvement

www.oecd.org/edu/evaluationpolicy

Consult this publication on line at ***http://dx.doi.org/10.1787/9789264190658-en.***

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases.
Visit ***www.oecd-ilibrary.org*** for more information.

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-19064-1
91 2013 02 1 P

