OECD Education Working Papers No. 58

Integrating Formative and Summative Assessment: Progress Toward a Seamless System?

Janet Looney

https://dx.doi.org/10.1787/5kghx3kbl734-en

OECD

Unclassified **EDU/WKP(2011)4**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **08-Apr-2011**

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2011)4
Unclassified

Cancels & replaces the same document of 10 February 2011

INTEGRATING FORMATIVE AND SUMMATIVE ASSESSMENT: PROGRESS TOWARD A SEAMLESS SYSTEM?

OECD Education Working Paper No. 58

by Janet W. Looney

This paper was commissioned to Janet Looney, an independent consultant specialising in programme design, evaluation and learning. The paper forms part of the work undertaken by the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes and includes revisions in light of the discussion of an earlier version [EDU/EDPC/EA(2010)2] at the 2nd meeting of the Group of National Experts on Evaluation and Assessment (9-10 September 2010).

The OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes is designed to respond to the strong interest in evaluation and assessment issues evident at national and international levels. The overall purpose is to explore how systems of evaluation and assessment can be used to improve the quality, equity and efficiency of school education. The Review looks 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 assessment and system evaluation. More information is available at: www.oecd.org/edu/evaluationpolicy.

Contact: Mr. Paulo Santiago [Tel: +33(0) 1 45 24 84 19; e-mail: paulo.santiago@oecd.org]
and Ms. Deborah Nusche [Tel: +33(0) 1 45 24 78 01; e-mail: deborah.nusche@oecd.org].

JT03299965

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

OECD DIRECTORATE FOR EDUCATION

OECD EDUCATION WORKING PAPERS SERIES

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Directorate for Education. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

Comment on the series is welcome, and should be sent to either edu.contact@oecd.org or the Directorate for Education, 2 rue André Pascal, 75775 Paris CEDEX 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

Applications for permission to reproduce or translate all, or part of, this material should be sent to OECD Publishing, rights@oecd.org or by fax 33 1 45 24 99 30.

Copyright OECD 2011

ABSTRACT

A long-held ambition for many educators and assessment experts has been to integrate summative and formative assessments so that data from external assessments used for system monitoring may also be used to shape teaching and learning in classrooms. In turn, classroom-based assessments may provide valuable data for decision makers at school and system levels. Currently there are important technical barriers to this kind of seamless integration. Nevertheless there are a number of promising developments in the field. Ongoing research and development aims at improving testing and measurement technologies, as well strengthening classroom-based formative assessment practices. Improved integration of formative and summative assessment will require investments in new testing technologies, teacher training and professional development, and further research and development.[1]

RÉSUMÉ

L'intégration des évaluations sommative et formative des élèves a toujours été une ambition des éducateurs et des experts afin d'assurer que les données utilisées pour le monitoring des systèmes d'éducation puissent également servir pour améliorer les processus d'apprentissage dans les salles de classe. En retour, l'évaluation des élèves en salle de classe peut fournir des données précieuses pour les décideurs aux niveaux de l'école et du système d'éducation. Actuellement, il y a des obstacles techniques importants à la réalisation de cette intégration des évaluations sommative et formative. Néanmoins, certains développements prometteurs dans ce domaine ont vu le jour. Les travaux de recherche et développement essayent aujourd'hui d'améliorer les techniques de tests et de mesure et de renforcer les pratiques d'évaluation formative en salle de classe. Une meilleure intégration des évaluations formative et sommative des élèves nécessitera des investissements dans de nouvelles technologies de tests, dans la formation des enseignants et dans la recherche et développement.

[1] ***Janet Looney***, an American national, is an independent consultant specialising in programme design, evaluation, and learning. Between 2002 and 2008, Ms. Looney was the project lead for the What Works in Innovation in Education programme at the OECD's Centre for Educational Research (CERI). She led the development of two major international synthesis reports: *Formative Assessment: Improving Learning in Secondary Classrooms* (2005), and *Teaching, Learning and Assessment for Adults: Improving Foundation Skills* (2008). Prior to her work with the OECD, Ms. Looney was Assistant Director of the Institute for Public Policy and Management at the University of Washington (1996-2002), where she was involved in evaluation of community development programmes, urban education reforms, and state-level implementation of federal welfare. Between 1994 and 1996, she was a Programme Examiner in the Education Branch of the U.S. Office of Management and Budget. She received her Master of Public Administration and Master of Arts in International Studies degrees from the University of Washington in 1993.

TABLE OF CONTENTS

SECTION 1: INTRODUCTION 5

SECTION 2: WHAT IS FORMATIVE ASSESSMENT? 7

2.1 What is the impact of formative assessment on teaching and learning? 7
2.2 The elements of formative assessment 8
2.3 Putting formative assessment into practice 10

SECTION 3: OVERVIEW OF POLICY APPROACHES 11

3.1 An emphasis on accountability 11
3.2 Assessment for school and system level improvement 12
3.3 Policies supporting formative assessment 13

SECTION 4: LINKING LARGE-SCALE, STANDARDS-BASED ASSESSMENTS AND CLASSROOM-BASED FORMATIVE ASSESSMENT 15

4.1 Uneven progress across the disciplines of cognitive science and educational measurement 16
4.2 Timing: long-, medium- and short-cycle formative assessment 17
4.3 The role of stakes 18
4.4 Performance-based assessments 19

SECTION 5: TEACHER APPRAISAL 22

SECTION 6: STRENGTHENING THE LINKS BETWEEN LARGE-SCALE, STANDARDS-BASED ASSESSMENTS AND CLASSROOM-BASED FORMATIVE ASSESSMENT 24

6.1 Strengthen teachers' assessment roles 24
6.2 Strengthen Teacher Appraisal 25
6.3 Draw on advances in cognitive sciences to strengthen both formative and summative assessment 25
6.4 Develop curriculum-embedded or "on-demand" assessments 26
6.5 Use diagnostic assessments for students at lower proficiency levels to better identify specific learning needs 27
6.6 Consider population sampling for large-scale assessments used for monitoring purposes 27
6.7 Take advantage of technology 28

SECTION 7: GENERAL POLICY IMPLICATIONS AND CONCLUSIONS 29

7.1 Learn from the bottom up: use formative assessment data to build knowledge about what works in policy and practice 29
7.2 Promote teacher professionalism 29
7.3 Ensure cost effectiveness by developing more effective approaches to assessment 29
7.4 Address Gaps in Research and Development 30

REFERENCES 31

ANNEX 1: ASSESSMENT AND EVALUATION FRAMEWORKS OECD COUNTRY POLICIES 37

ANNEX 2: CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE) 52

ANNEX 3: OECD COUNTRY POLICIES ON ASSESSMENT OF TEACHER PERFORMANCE 62

SECTION 1: INTRODUCTION

1. Student assessment has taken an increasingly prominent role in education policy in OECD countries. As the majority of OECD countries have decentralised education systems so that schools may better shape provision to meet local needs, many countries and regions have also developed large-scale assessments to monitor student and school performance. Schools are held accountable for helping students to meet central standards, as measured by these national or regional assessments. Policy makers and school leaders also use the assessment data to identify strengths and weaknesses in student and school performance, and to improve the quality of teaching and learning.

2. Classroom-based "formative assessment" has also taken on an increasingly important role in education policy in recent years. Formative assessment refers to the frequent, interactive assessment of student progress to identify learning needs and shape teaching (OECD, 2005). Black and Wiliam's 1998 review of rigorous quantitative studies established that formative assessment methods and techniques produce significant learning gains – according to their analysis, among the largest ever identified for educational interventions. Moreover, a few studies have shown the largest gains for students who had previously been classified as low achievers.

3. Formative assessment, which emphasises the importance of actively engaging students in their own learning processes, resonates with countries' goals for the development of students' higher-order thinking skills and skills for learning-to-learn. It also fits well with countries' emphases on the use of assessment and evaluation data to shape improvements in teaching and learning.

4. A long-held ambition for many educators and assessment experts has been to integrate summative and formative assessment more closely so that data from external assessments used for system monitoring may also be used to shape teaching and learning in classrooms, and in turn, classroom-based assessments may provide valuable data for decision makers at school and system levels[2]. Currently, however, there are important technical barriers to this kind of seamless integration. Typically, data gathered in large-scale assessments are not at the level of detail needed to diagnose individual student needs, nor are they delivered in a timely enough manner to have an impact on the learning of students tested. There are also challenges related to creating reliable measures of higher-order skills emphasised in standards and curricula, such as problem solving and collaboration.

5. High stakes associated with external assessments, such as the threat of school reconstitution or shut down are intended to focus teachers' attention on educational standards and priorities, but they may also undermine innovative approaches to teaching, including formative assessment. There is evidence that teachers are more likely to "teach to the test" when assessments are perceived as having high stakes. At the same time, OECD countries have paid scant attention to the role of teacher appraisal as a means for monitoring the quality and impact of teaching and classroom-based assessment. As a result, there have been few efforts to develop valid measures of teachers' teaching and assessment practices (Herman *et al.*,

[2] The report does not cover research on tests or examinations that are used for selection purposes (*i.e.* for admission to programmes or higher education) at any length, because these tests results are not typically used formatively. The report does not cover targeted evaluations of innovative educational projects, as these are often ad hoc rather than systematic evaluations.

2010), and missed opportunities to provide teachers with formative feedback on their own performance and to reinforce innovative practices.

6. While acknowledging some of the limits of current assessment technologies and practices, the overall message of this report is very positive, as there are a number of promising developments in the field. These include efforts to develop more coherent and coordinated assessment and evaluation frameworks. There is also ongoing research and development aimed at improving testing and measurement technologies – several of which are also aimed at improving classroom-based formative assessment practices.

7. The following section (Section 2) provides an overview of international research on formative assessment and evidence of its impact on student learning. It describes the elements of effective classroom-based formative assessment, and provides a foundation for understanding policy and school environments that support successful practice.

8. Section 3 provides an overview of broader assessment frameworks that are part of standards-based frameworks in OECD countries. While systems share many key features – combining external assessments with support for internal, classroom-based assessment and school self-evaluations – there are also variations in design and approach. Different OECD countries use a variety of policy levers to promote and support classroom-based formative assessment. This overview, along with the discussion in Section 2, helps to set the context for the subsequent sections.

9. Section 4 is, in many ways, at the core of this report. The focus is on some of the technical barriers to closer integration of classroom-based formative assessment with large-scale, standards-based assessments. Close examination of current barriers is vital for development of new assessment technologies. The fifth section briefly examines how teacher appraisal might support more effective and systematic practice of classroom-based formative assessment, while the sixth section focuses on approaches to strengthening the links between large-scale, standards-based assessments and classroom-based formative assessments.

10. Section 7 concludes the report. It sets out broad policy implications of the discussion, and proposals for stronger integration of formative and summative assessments, with the ultimate goal of improving student achievement.

SECTION 2: WHAT IS FORMATIVE ASSESSMENT?

11. The concepts of "formative" and "summative" assessment are, of course, central to this report[3]. Summative assessment refers to summary assessments of student performance – including tests and examinations and end-of-year marks. Summative assessments of individual students may be used for promotion, certification or admission to higher levels of education. Formative assessment, by contrast, draws on information gathered in the assessment process to identify learning needs and adjust teaching. Summative assessment is sometimes referred to as assessment *of* learning, and formative assessment, as assessment *for* learning.

12. Scriven (1967) first suggested the distinction between formative and summative approaches in reference to evaluations of curriculum and teaching methods. He suggested that evaluators could gather information early in the process of implementation to identify areas for improvement and adaptation, and at successive stages of development. Soon after, Bloom (1968) and Bloom, Hasting and Madaus (1971) took up this idea, applying the concept to student assessment in their work on "mastery learning". They initially proposed that instruction be broken down into successive phases and students be given a formative assessment at the end of each of these phases[4]. Teachers would then use the assessment results to provide feedback to students on gaps between their performance and the "mastery" level, and to adjust their own teaching to better meet identified learning needs (Allal, 2005).

2.1 What is the impact of formative assessment on teaching and learning?

13. Since this early work on formative assessment and evaluation, researchers working in different linguistic traditions have contributed to a wide-ranging literature aimed at both refining and enlarging the concept (see Allal and Mottier-Lopez and Köller reviews of the French- and German-language literature on formative assessment, both included in OECD, 2005). Formative assessment is now seen as an integrated part of the teaching and learning process, rather than as a separate activity occurring after a phase of teaching (Allal, 1979, 1988; Audibert, 1980; Perrenoud, 1998). It encompasses classroom interactions, questioning, structured classroom activities, and feedback aimed at helping students to close learning gaps. Students are also actively involved in the assessment process through self- and peer-assessment (Sadler, 1989). Information from external tests or from school inspections may also be used formatively to identify learning needs and adjust teaching strategies. The crucial distinction is that the assessment is formative if and only if it shapes subsequent learning (Black and Wiliam, 1998; Wiliam, 2006).

[3] Much of the information on countries' assessment and evaluation policies was gathered from UNESCO's World Data on Education database, which provides a systematic overview. In describing assessment policies, several country reports use the term "continuous assessment" or "ongoing assessment" to refer to frequent assessment of student progress (which may refer to both formative and summative assessments). However, the reports do not provide information on country or regional policies to promote these classroom-based assessments.

[4] The concept of mastery learning draws on Vygotsky's "zone of proximal development" (ZPD). The ZPD is the difference between what the student is able to do with help and what he or she can do without guidance. As a student progresses toward mastery, he/she student gradually becomes more independent. This is also a key concept in formative assessment (Griffin, 2007).

14. In their seminal review of the research on classroom-based formative assessment, Black and Wiliam (1998) studied the impact of different approaches and techniques on student learning[5]. Their review draws on 250 international sources, covering 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, such as mastery learning (in which, as noted above, the concept of student formative assessment has its origins); and, student self- and peer-assessment.

15. Drawing upon the evidence gathered for the review, Black and Wiliam concluded that the achievement gains associated with formative assessment were among the largest ever reported for educational interventions, and if replicated across a countries, would "increase in the score of an "average" ranking, as measured by the international Trends in Mathematics and Science Study (TIMSS) to ranking among the top five countries".

16. The Black and Wiliam 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. Several OECD countries now promote formative assessment as a key strategy for meeting goals for quality and equity (see Section 3).

2.2 The elements of formative assessment

17. 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. Initial work on formative assessment changed this approach somewhat by incorporating tests *within* study units, for example, when students had finished working on a specific learning activity, in order to allow teachers to diagnose learning needs and adjust teaching at that point. The assessments were nevertheless still seen as being separate from normal classroom activities.

18. In the early 1980s, Audibert suggested that formative assessment might be incorporated into daily teaching activities, allowing teachers and students to adapt teaching and learning on an ongoing basis. ***Formative assessment is thus seen as an integrated part of teaching, learning and assessment***. Audibert proposed that this approach would allow students to engage in conscious reflection of the learning process, as well.

19. ***Classroom cultures*** are also important to effective formative assessment practice. They encompass relationships between and among students and teachers, as well as beliefs about learning and learners. As Shepard and colleagues (2005) caution, adopting the techniques of formative assessment without any corresponding shift in philosophy is likely to undermine efforts. Similarly, students need to develop new understandings of themselves as learners.

20. A key issue that emerged in the OECD's (2005) international study on formative assessment as practiced in exemplary classrooms was the importance of helping 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.

[5] Earlier reviews by Natriello (1987) and Crooks (1988) reached substantially the same conclusions as the 1998 Black and Wiliam review. Black and Wiliam (2003) suggest that their 1998 review may have had a larger impact than previous reviews as a result of outreach efforts – through publication of a short guide for practitioners, *Working Inside the Black Box* (Black *et al.*, 2002), as well as through active media dissemination.

21. Several studies have shown 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. It is also important to "scaffold" information given in feedback – that is, to provide as much or as little information as the student needs to reach the next level. Feedback that is non-specific (*e.g.* "needs more work") or "ego-involving", even in the form of praise, may have a negative impact on learning (see for example, Boulet *et al.*, 1990; Butler, 1988). On the other hand, feedback that provides guidance on how to improve performance has a positive impact on learning.

22. Feedback focused on the learning process rather than the final product, and which tracks progress over time, has also been found to be more effective. Mischo and Rheinberg (1995) and Köller (2001) have identified several experimental studies where teachers tracked progress over time, showing positive effects on students' intrinsic motivation, academic self-concept, performance, and attribution of achievement to effort as opposed to ability. Findings from OECD's Programme for International Student Assessment (PISA) reinforce this research. PISA 2000, which focused on reading literacy of 15 year olds, found that students who had learned to manage their own learning processes tended to perform better on the PISA reading literacy scale (OECD, 2001).

23. Other studies focus on the timing of feedback. Feedback is most effective when it is provided within minutes (or even seconds) – or at the most, within a period of days (Wiliam, 2006). At the same time, feedback should not be provided too rapidly – *i.e.* before the student has had a chance to try to work out a problem him or herself.

24. Effective ***questioning*** techniques help to reveal students' level of understanding and identify possible misconceptions (in contrast to questions that are designed to elicit a "yes" or "no" response or that stress recall rather than reasoning processes provide little information on the student's level of understanding and may hide errors in thinking). Questions may explore students' understanding regarding the direction of causality in a process they are just learning about, or "why" questions, will help to reveal possible misconceptions. Teachers may also guide students toward deeper understanding of a subject through extended dialogues that build on a series of questions (OECD, 2005). Students may develop and deepen knowledge by generating their own lines of questioning (Williams and Ryan, 2000).

25. Teachers may also gain insight into student thinking through observation, review of written work products and portfolios, student presentations and projects, interviews, tests and quizzes (Shepard, 2006). These varied views on student work over time and in different contexts allow teachers to identify patterns in thinking and problem solving.

26. A fundamental goal for formative assessment is to help students develop skills for ***self- and peer-assessment*** (Sadler, 1989). Teachers establish clear learning goals and share criteria for assessing the quality of work with students. Students thus develop skills to monitor their own work so they can gauge how well they are doing in relation to a set standard. They may develop new understandings of who they are as learners, and strengthen self-efficacy (belief in the ability to accomplish specific tasks). Again, the focus is on the process of learning as much as it is on the outcome. Students build skills for "learning to learn".

27. The OECD (2005) study on formative assessment practice in exemplary classrooms found that teachers drew on each of the different elements explored above in some measure and that the elements were mutually reinforcing. Teachers in the OECD study also noted the importance of being more systematic in their approach to classroom assessment, as the most effective interactions with students are the result of careful planning.

28. Formative assessment is thus seen as an integrated part of the teaching and learning process. Effective practices are grounded in theories about learning and performance (cognition) in a given subject domain. Teachers establish goals that are appropriate to learners' development level and create learning situations that will help students to grasp new concepts. They may also develop questions or activities that may reveal misconceptions (Black, 2000). The process is iterative. Over time, students acquire new knowledge and create new, increasingly coherent mental frameworks for understanding. New types of evidence of student progress and understanding are needed at successive stages.

29. It is also important to note that approaches to teaching, learning and assessment need to be adapted to the domain being studied. For example, students learning to read must develop and draw upon a range of skills, all of which are used simultaneously. These include phonological awareness, decoding, vocabulary, knowledge of grammar and language structures and reasoning skills. If students have difficulty, teachers need to assess and explore a range of potential causes in order to develop an appropriate teaching intervention. In mathematics, teachers must assess and explore student's grasp of basic concepts where they may have some difficulty, as well as their computational skills before they are able to adapt teaching (Honig, 2001).

2.3 Putting formative assessment into practice

30. Several OECD countries have developed policies to support formative assessment practice (explored in more detail in the section below). However, while evaluations of specific pilot programmes to build teachers' formative assessment capabilities have been positive (Wiliam *et al*., 2004), there are no system-wide evaluations of the impact of these policies on teaching practice or student achievement.

31. According to some studies, effective implementation of formative assessment may be more the exception than the rule (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 & Cowie, 2001; Heritage, 2010; Herman *et al.,* 2010). But it is much more typical to find that teachers emphasise rote learning, develop only superficial questions to probe student learning, and provide only general feedback. 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 protest that that there are too many 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, 2005).

32. There is also significant evidence that external assessments and evaluations, particularly in systems that attach high stakes to results, encourage teachers to "teach to the test". Poor alignment between external assessment and evaluation and classroom assessments may also undermine practice. These issues are explored in more detail in subsequent sections.

SECTION 3: OVERVIEW OF POLICY APPROACHES

33. At the policy level, assessments and evaluations have been developed to meet a range of purposes in OECD countries. Among these aims are:

- Accountability – Data on educational performance are made available to taxpayers, parents and policy makers, who want to know whether schools are meeting standards. In systems that promote school choice and competition, these data may also support parent and student decisions as to where they will find the best education for their needs. Accountability is also seen as a way to motivate improvement.

- School and system improvement – School leaders, teachers and policy makers may refer to data on school and student performance 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 so on. Teachers may also use the data to shape general teaching strategies. This is essentially formative use of data.

- Support for student learning through classroom-based formative assessment – Information on individual student progress and understanding is used to adapt teaching. The focus is on helping all students meet learning gaps.

34. How systems can most effectively balance these different aims for assessment is the subject of much debate.

35. This section provides a brief overview of assessment and evaluation frameworks in different OECD countries. It is important to remember that these policies are continuously evolving. Moreover, within a single country it is possible to find very different approaches to assessment and evaluation in different regions. The impact of policies will also depend on the details of programme design and implementation. While keeping these caveats in mind, the different country approaches may provide a rich laboratory for learning.

3.1 An emphasis on accountability

36. Generally, countries emphasise either the accountability or improvement functions of external assessment and evaluation. While both approaches are focused on improvement – they reflect fundamentally different ideas as to how to motivate change.

37. Countries that place greater emphasis on accountability may attach high stakes to school and student performance as measured in assessments and evaluations. A relatively small number of countries and regions fall within this category – they include Canada, the United Kingdom and the United States. Stakes may include teacher job loss, school reconstitution or shut down. The idea is that high stakes will provide incentives for both teachers and students to work harder and more effectively. Schools use information from assessment and evaluation to identify weak areas, and to reallocate resources and/or to develop new instructional strategies (Jacob, 2003). At the same time, high stakes have been shown to lead to narrowing of curriculum, and score inflation as teachers “teach to the test” to avoid sanctions. (More will be said about the role of stakes in Section 4.)

38. Studies have shown that some teachers behave as if assessments have high stakes even when the results are used only for improvement. According to the OECD's (2008) *Education at a Glance* (EAG), publication of assessment results, as occurs in the majority of OECD countries, adds to the stakes. Teachers will work to avoid the stigma of a low rating (Corbett and Wilson, 1991; Madaus, 1988; McDonnell and Choisser, 1997). At least 18 OECD countries[6] publish the results of external assessments and/or evaluations (inspections and/or school self-evaluations). They include Belgium (the Flemish Community), the Czech Republic, Denmark, England, France, Hungary, Iceland, Korea, the Netherlands, New Zealand, Norway, Portugal, Scotland, Slovenia, Sweden and Turkey. According to *Education at a Glance*, Australia, Ireland and Italy also publish results, but avoid the use of tables that compare school performance. In the Flemish Community of Belgium, policy makers have taken the unusual measure of legally forbidding publication the results on a comparative basis. Only a few countries avoid publication of the results of external student assessments and/or school evaluations altogether – thus avoiding associated stakes. They include Finland, Mexico and Luxembourg[7].

39. It is also worth noting that international assessments, such as the Trends in Mathematics and Science Survey and the OECD's Programme for International Student Assessment (PISA) have influenced country decisions to introduce external assessments, for example in Denmark and the German Länder, where there previously had been little emphasis on external monitoring.

40. Technically, school-leaving examinations are beyond the scope of this report because results at this final phase of upper-secondary school are not used formatively to identify learning needs of individual students. However, these examinations *do* have some impact on teaching and learning, as teachers may adapt teaching for future groups of students in areas where the graduating cohort performed poorly. In countries offering school choice, there are also stakes attached. Parents and students identify the "best" schools as those with high scores on school leaving examinations, as well as admission to prestigious tertiary institutions.

41. School-leaving examinations are the primary form of large-scale student assessment in fewer than one-third of OECD countries (Austria, the Czech Republic, Hungary, and Slovakia). Other countries administer school-leaving examinations in addition to large-scale assessments for accountability and monitoring (Denmark, Finland, France, Korea, Luxembourg, Italy, the Netherlands, Norway, Poland, Portugal, Sweden the United Kingdom and several states in the US).

3.2 Assessment for school and system level improvement

42. Several countries place greater emphasis on use of assessment and evaluation results for school and system level improvement. The stakes associated with the results of assessment are relatively low. Rather, the emphasis is solely on use of the information gathered in assessment and evaluation as a means to improving performance. It is essentially a formative approach.

43. In the French-speaking Community of Belgium, France and Spain large-scale, external assessments are considered as diagnostic. They are administered at the beginning of new phases in schooling, such as the transition from primary to lower secondary school. The aggregated data are used to identify categories of student needs and to develop appropriate policies.

[6] Data on publication of assessment results not available or not applicable for the remaining 13 countries and regions.

[7] The French-speaking community of Belgium and Poland administer large-scale assessments of students in selected years, but have not provided information on publication or the results. The German Länder are now developing assessments for the purpose of monitoring. There is no information on publication of results.

44. Some countries administer assessments to only a sample of the student population (this is referred to as population sampling, while assessments that are administered to all students are referred to as census testing). In this way, it is possible to track trends in student performance across different demographic groups, and to develop appropriate policies. According to UNESCO's World Data on Education database, Canada, Finland, Korea and the US take this approach. However, in Canada and the US this approach is used only with national assessments; at the province/state level, assessments are given to every student at selected year levels.

45. Another approach to lowering stakes associated with school-leaving examinations is to combine results with teacher's classroom-based assessments and observations. Denmark, Greece, the Netherlands, Norway, Poland, Sweden, Switzerland and the UK take this approach. In Queensland, Australia, education authorities eliminated standardised external assessments in 1972 and introduced a system of teacher-moderated assessment of student portfolios (OECD, 2005).

3.3 Policies supporting formative assessment

46. Several countries and regions also provide policy support for classroom-based formative assessment (see Annex 2). The OECD's study on formative assessment in lower secondary schools provides the most systematic overview of different country polices on formative assessment currently available[8] (OECD, 2005). The policies are aimed at building teachers' and school leaders' assessment skills, creating opportunities for teachers to innovate, and providing guidelines and tools to facilitate formative assessment practice. For example, legislation governing the Danish *folkeskole* system requires schools to use student assessment as the basis for student guidance and to shape teaching methods. Italy requires teachers to use a "valuation form" to track students' learning and development (including social, behavioural, cognitive and metacognitive) and to facilitate communication between students, parents and teachers.

47. Several countries have developed curriculum guidelines to assist teachers in more systematic integration of formative assessment. In 2000, England introduced the Assessment for Learning (AfL) programme in lower secondary schools (Key Stage 3). Scotland's own Assessment is for Learning (AiFL) programme similarly encourages teachers to consider assessment as an integrated part of teaching and learning process. The Department of Education in Newfoundland and Labrador, Canada, disseminates rubrics with specific guidelines and criteria for evaluating student work.

48. New Zealand first introduced its National Assessment Strategy in 1999, providing assessment tools and professional learning to build assessment capabilities. Since then, the strategy has evolved and expanded. New assessment tools were introduced through the asTTle (assessment tools for teaching and learning, now available in a 4th version), the government has published exemplars focusing on curriculum and formative assessment principles. Most recently, the National Education Monitoring Programme (NEMP) has been modified to include information useful to implementation of the new *National Standards*, which are based on "assessment for learning" principles.

49. In New Zealand, primary level student assessments are based on teachers' qualitative judgments, of student performance and progress. There are no national tests. At the secondary level, the National Certificate of Educational Achievement (NCEA) sets standards for student performance. Students are evaluated against written criteria, which are accompanied by exemplars showing expected levels of student

[8] Note that there is no systematic overview of policies to support formative assessment practice across all OECD countries at this point. The 2005 OECD study covered formative assessment policies in: Canada, Denmark, England, Finland, Italy, Scotland, New Zealand, and the state of Queensland in Australia.

performance. Since 2008, Māori assessment experts have been developing assessment tools to be used in Māori-medium settings.

50. These different country policies fit within and are affected by broader frameworks for assessment and evaluation. The next section explores a range of challenges to integrating large-scale, standards-based assessments and classroom-based formative assessments. The discussion will then turn to potential approaches to addressing these challenges and to improving integration of large-scale summative and classroom-based formative assessments.

SECTION 4: LINKING LARGE-SCALE, STANDARDS-BASED ASSESSMENTS AND CLASSROOM-BASED FORMATIVE ASSESSMENT

51. Large-scale assessments provide useful data for monitoring overall performance of education systems and of individual schools and groups of students. As has been noted, the data help shape decisions on educational policy directions, curriculum needs, allocation of financial resources, as well as adaptation of general instructional strategies. These assessments also help to keep schools focused on student achievement, and reinforce national or regional educational standards.

52. It is sometimes assumed that data gathered in large-scale, standards-based assessments might also be used to create profiles of individual students' learning needs. But, there are real limits to the extent to which data from large-scale, standards-based assessments may be used to target specific student needs or to shape classroom instruction:

- While there have been important advances in the cognitive sciences – that is, the understanding of how students learn – large-scale assessments, which are designed to ensure that data are valid, reliable and generalisable, cannot easily capture student performance on more complex tasks, such as problem solving, reasoning, or collaborative work. These large-scale, standards-based assessments do not provide the detailed information needed to diagnose the specific sources of student difficulty.

- Feedback from large-scale, standards-based assessments is usually delivered to schools several weeks after tests are administered (recall research cited above on the need to provide formative feedback in a timely manner).

- While high-stakes assessments focus teachers' attention on helping students to meet central standards, there is evidence that many teachers also narrow instruction, focusing attention on those areas most likely to be tested. When this occurs, tests no longer serve as proxies of wider achievement. Scores overstate students' performance, and fail to provide accurate information on student progress.

- Performance-based assessments may address some of the problems associated with large-scale, standardised assessments. However, they are more costly to design, administer and score. There are also challenges in ensuring that scores for these kinds of assessments are both reliable and generalisable.

53. In spite of these challenges, there are several potential strategies to help strengthen the correspondence between the different levels of assessment, so that results may be used to shape improvements at every level of the system. Moreover, ongoing research aimed at strengthening large-scale, standards-based assessments and addressing some of the shortcomings described in more detail below will also help to support more effective classroom-based formative assessment. Several of these possibilities will be discussed at the end of this section.

4.1 Uneven progress across the disciplines of cognitive science and educational measurement

54. Over the past several decades – and in particular since the early 1990s – cognitive scientists have made a great deal of progress in understanding the process of learning in different subject domains. This includes understanding novice performance and typical learner misconceptions, the development of effective learning environments, and the importance of developing students' capacity to monitor their own learning and to assess the effectiveness of their learning strategies (self-assessment and metacognitive monitoring) (Bransford *et al.*, 1999; Pellegrino *et al.*, 1999).

55. Domain specific research has also yielded important information on learning processes. For example, research on the "psychology of mathematics education" explores the ways in which students understand mathematics curriculum and common errors in responses. Teachers can therefore better anticipate the kinds of misunderstandings students are likely to have and to plan instruction accordingly. They may also analyse patterns in student responses to questions, tracking the different ways that learners may take in and understand new information (Harlen and James, 1997; Williams and Ryan, 2000).

56. However, 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 (Chudowsky and Pellegrino, 2001; Gipps, 1996; Mislevy, 1993; Pellegrino *et al.*, 1999). This has been particularly true for large-scale, standards-based assessments (whether based on traditional standardised, multiple choice or tests using alternative formats).

- ***A first challenge*** is related to the difficulty of deconstructing cognitive performances for purposes of measurement. In traditional testing methodology, tasks are treated as discrete items and student responses to different tasks are aggregated as an overall score. However, this methodology is at odds with research emphasising learning as the "continuous acquisition and restructuring of domain-based knowledge". Expertise involves both declarative and procedural knowledge (learning not only *what* but also *how to*) (Gipps, 1996; Pellegrino *et al.*, 1999, p 317).

- ***A second challenge*** is related to how student scores are reported. Assessment results are typically reported as either "norm-referenced" (*i.e.* describing student performance relative to his/her peers), or "criterion-referenced" (*i.e.* describing student performance relative to a performance target). Several measurement experts argue that, while norm-referenced assessments are useful for the purpose of selection (*e.g.* for school or university admissions), criterion-referenced assessments are more instructionally useful because they measure student progress toward specific goals – and this approach is in line with the formative assessment focus on helping all students to close learning gaps and meet goals.

  In criterion-referenced systems, scores (which may be based on multiple choice, rubric or open-response items) are converted into a scale, which are then tied to broad proficiency categories, such as: below basic, basic, proficient, advanced (McGehee and Griffith, 2001). But several measurement experts argue that these categories are too broad to provide any kind of diagnostic information necessary for profiling individual student needs. Rupp and Lesaux (2006), for example, conducted an analysis of the relationship between student performance on a criterion-referenced, standards-based assessment of reading comprehension of fourth year students, and the performance on a diagnostic battery of component reading skills for a cohort of children

followed from pre-primary through the fourth year of school[9]. They found that the standards-based assessment provided only weak diagnostic information, and masked significant heterogeneity in the causes of poor performance. In order to identify the cause of poor performance and develop an appropriate instructional intervention, teachers needed to administer additional assessments with greater diagnostic precision. Similarly, Buly and Valencia (2002) found that teachers in the US state of Washington developed remediation plans for fourth year students who had performed poorly on the state's standardised assessment of reading skills by providing additional phonics instruction, which was appropriate for only some of these students.

- ***A third challenge*** is related to the difficulty of balancing technical concerns for the generalisability (the results of a test can be generalised to other tests or groups), reliability (the test can be repeated and produces consistent scores) and validity of assessment data (the test measures what it is intended to measure). Generalisability and reliability are of particular importance in the context of large-scale, standards-based assessments, as performances are compared across large numbers of schools and students. Validity issues, on the other hand, are much more likely to be the key concern for classroom-based assessments. Within the classroom context, the validity of assessments are based on connections between the learning goal being assessed, the questions or tasks being used to gauge student understanding, and the way in which teachers interpret and act upon student responses to close any learning gaps. Questions or tasks need to yield appropriate inferences and with sufficient detail in order to guide subsequent instruction (Herman, *et al.*, 2010).

57. If systems are to integrate large-scale, standards based assessments and classroom-based formative assessments, they will need to find a better balance both within as well as across these different approaches. As Pellegrino and colleagues (1999) observe, each approach has specific limitations. They note that by "...selectively focusing on a specific assessment purpose (summative vs. formative) as applied to a specific assessment context (large scale and high stakes vs. classroom based and low stakes), one or more critical issues of inference are largely ignored" (p. 332).

4.2 Timing: long-, medium- and short-cycle formative assessment

58. Several researchers distinguish different levels of formative assessment based on timing and purpose. Allal and Schwartz (1996) refer to formative assessments that directly benefit students who are assessed as "Level 1", and formative assessments where data gathered are used to benefit future instructional activities or new groups of students as "Level 2". Alternatively, Wiliam (2006) distinguishes between long-, medium-, and short-cycle formative assessment. According to Wiliam's definition, long-cycle formative assessment occurs across marking periods, semesters or even years (four weeks to a year or more). A 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). Shavelson *et al.* (2008) refer to the rapid feedback based on exchanges between teachers and students, or between peers as "on-the-fly" formative assessment.

[9] Rupp and Lesaux note three important differences between standards-based and diagnostic assessments: 1) Diagnostic measures of reading are based on a well-established, large body of research related to the different components of the reading process. There is much less accessible empirical evidence on the construct validity of standards-based assessments of reading comprehension; 2) Diagnostic measures of the component skills of reading comprehension are usually administered to individual students. Standards-based assessments are given to groups of students; 3) Diagnostic assessments provide a profile of the individual student's component skills in reading. Standards-based assessments measure composite reading skills and are reported within broad proficiency classifications.

59. Assessment data appear to have the most impact on student achievement when delivered in timely manner. Data from large-scale, standards-based assessments, however, are usually available to teachers several weeks to months following the actual test day. 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), the impact on student achievement appears to be modest. By contrast, Wiliam and colleagues (2004) found that in classrooms where teachers provided formative feedback within or between teaching units – for instance, during an in-class interaction or over the course of a month-long teaching unit – the rate of student progress over the year was approximately double that found in the control classrooms.

4.3 The role of stakes

60. Countries with a strong emphasis on school accountability, as noted in the overview, are more likely to attach high stakes to results of external assessments. Stakes are intended to focus attention on priorities of national standards and/or curricula. Data from the large-scale, standards-based assessments are intended to provide a clear picture of how students in a particular school or class are performing. If a student or group of students fails to meet standards, teachers and school leaders will search for more effective instructional methods. Thus, high-stakes, large-scale assessments are expected to have a somewhat indirect, although important impact on improving the quality of teaching and learning.

61. Educational measurement experts warn that high-stakes assessments may also have a number of unintended consequences. High stakes may create incentives to "teach to the test". Teachers may coach students on test taking strategies and tricks (*i.e.* non-substantive aspects of tests), or re-align focus on the content and kinds of problems most likely to appear on test, based on patterns identified in tests over past year (*i.e.* substantive aspects of tests), or re-allocate time spent on higher priorities subjects. Teachers significantly narrow learning if they focus only on content and skills that are most likely to be on a test, since no single test can measure the full range of skills and knowledge set out in standards and curriculum. Teachers may also be more likely to focus on rote learning and memorisation of superficial facts, rather than higher-order skills.

62. Re-allocation, realignment and coaching can lead to test score inflation. In other words, test results will overstate the students' actual learning. Tests may also include a significant level of error. For example, students may misunderstand the question or the problem being posed (Gauld, 1980) and therefore answer incorrectly. Both score inflation and error rates make it difficult for school leaders and teachers to interpret results and develop appropriate strategies for improvement.

63. The stakes for schools, teachers or individual student are of course higher when judgments of performance are based on a limited number of measures (*e.g.* a single high-visibility test or infrequent inspection visits). School leaders and teachers also have less information for identifying strengths and weaknesses, and planning for improvements (for a review of the impact of high-stakes assessments on educational innovation, see Looney, 2009).

64. Empirical evidence on the impact of high-stakes assessments on classroom instruction is mixed, although a number of studies report neutral or negative effects. Two macro-level studies on results of the National Assessment of Educational Progress (NAEP) in the US came to very different conclusions. Linn (2000) compared NAEP with state level assessments and found no clear trends, making it difficult to make any kind of generalisation about gains on state assessments. However, in a later study, Hanushek and Raymond (2005) found gains in student performance on the NAEP – with an effect size of 0.2 standard deviations. Because the gains were only in states attaching consequences to student performance on state-level assessments, the study claims support for the role of stakes in improving student achievement.

65. Other studies have focused more on evidence from the micro-level – that is, schools and classrooms. McDonnell and Choisser (1997) followed the implementation of assessment programmes in North Carolina and Kentucky, and found that teachers did make changes in instructional approaches, but most of these were relatively superficial. Two-studies on implementation of a high-stakes in reform in Chicago found mixed evidence of as to changes teachers made in their instructional practice. Abelmann, Elmore, Even, Kenyon, & Marshall (1999) found that teachers who had low expectations regarding their students' capacities, as well as their own capacity to influence learning, were less likely to change their instructional strategies in response to data from large-scale assessments. In his own review of the Chicago reform, Jacobs (2003) found that most improvements in student achievement were the result of increased student effort and parental involvement, re-alignment of teaching content. While increased effort is certainly a positive effect, it is notable that with very few exceptions, improvements were *not* linked to changes in instructional techniques, investments in teacher professional development, or reallocation of resources within schools.

66. Abrams and colleagues (2003) reviewed a survey of US teachers that had been conducted by the National Board on Educational Testing and Public Policy. The survey was focused on teachers' perceptions of the impact of the stakes associated with state assessments on teaching and learning. One of the most surprising findings, and most relevant to this report, was that a high percentage of teachers in both high- and low-stakes assessment environments, agreed that state-level assessments had a negative impact on their teaching. Seventy-six per cent of teachers working in high-stakes environments, and sixty-three per cent of those working in lower-stakes environments agreed that statewide assessments had led them to teach in ways that went against their own beliefs regarding effective practice.

67. Based on the evidence identified for this report, it appears that while high stakes may have an impact on the level of effort teachers, students and parents make, they have had very little effect on teacher's instructional strategies. The emphasis on large-scale assessment as a means to identify areas for improvement and adaptation of teaching has not necessarily led to adoption of similar strategies at the classroom level. Progress toward integration of large-scale, standards-based assessments and classroom-based formative assessments may help to bridge this gap.

4.4 Performance-based assessments

68. While there is currently limited information on the formats used for large-scale, standards-based assessments in different OECD countries, a few have provided basic information to the UNESCO World Data on Education (WDE) database.

- Canada implements the national School Achievement Indicators Programme (SAIP), sampling a small percentage of students across the country. Achievement is described over five levels, representing a continuum of competences. It includes multiple choice as well as short answer questions. There are also "practical" assessments of student's problem-solving skills in science, and communication skills in English.

- Denmark administers a computer-based, adaptive assessment[10]. Students who answer questions correctly are directed to a more difficult question, and those answering incorrectly are directed to an easier question. Since the test is adapted according to each student's responses, no two students take the same test, and it is not possible to compare student performances.

[10] No English-language studies reviewing the Danish assessment approach were identified for this report. However, computer-based, adaptive testing (CAT) is generally considered as providing more precise scores of student performance than typical standardised assessments.

- Korea implements the annual National Assessment of Scholastic Achievement (NASA) to a sample of 1% of all students at different school levels and across regions. NASA measures attainment of objectives in the school curriculum, and uses both multiple-choice and open response formats. For example, assessment of English and science subjects are based on student performances (*e.g.* demonstrated speaking skills in English; demonstration of processes used in science, or application of knowledge and skills to real-world problem).

- Sweden – National tests exist for key stages in compulsory school (Years 3, 5 and 9) and in upper secondary school. National assessments in Years 3 and 5 are intended for diagnostic and formative purposes. They are compulsory and must be administered by schools in a nationally specified period in the spring. The national tests in Year 9 and those in upper secondary school are summative. The results from national tests are one of the bases for teachers to determine students' overall grades. Teachers grade the national tests for their own students and each school decides how to weigh the national assessments and course grades (Nusche *et al.*, 2011).

- Under the No Child Left Behind Act (2000) in the US, each state develops its own assessment to track progress toward the state-level standards. Many states rely upon standardised, multiple-choice assessments. Several states have experimented with performance-based assessments (*e.g.* Vermont's statewide portfolio assessment programme, or Maryland's task-based performance assessments).

69. The strongest critiques of large-scale, standards-based assessments are usually directed at standardised tests that rely upon multiple-choice, close-ended question formats. While it is possible to develop tests using these formats that measure higher-order skills, it is not always easy to do. A number of alternative approaches to assessments have also been developed. These include performance-based assessments with open-ended prompts, exercises requiring written explanation, carrying out procedures, designing investigations, compiling a portfolio, making a performance, such as a speech or a musical recital. While standardised assessments are machine-scored, performance-based assessments are typically scored by human raters.

70. Performance-based assessments have also been seen as a way to shape teachers' approaches to instruction, ensuring that it is focused on development of higher-order skills, rather than rote memorisation. Several studies have shown that the performance-based assessments have had a positive impact on instruction– that is, teachers are more likely to adjust strategies so that they are in line with the tasks emphasised in the performance assessment. For example, Koretz, Stecher, Klein and McCaffrey (1994) reviewed implementation of Vermont's statewide portfolio assessment programme. They found that mathematics teachers reported they had increased their focus on mathematical problem solving and representation. In a 1998 study, Yoon and Resnick studied the implementation of the California Mathematics Renaissance (CMR) programme. The programme emphasised student group work, lab and fieldwork, oral presentations and portfolio development. All students, whether in the CMR programme or not, sat the performance-based New Standards Mathematics Reference Exam. The authors found that students in classrooms where teachers had reported using the kind of performance-based tasks emphasised in the CMR had higher scores on the examination[11]. On the other hand, Goldberg and Roswell (2000) in their study of the Maryland School Performance Assessment Program (MSPAP) found that teachers did not readily adjust instructional strategies. While the performance-based assessments may have, to some degree, facilitated teaching for higher order skills and formative assessment practice, the school districts across the state also needed to make significant investments in teacher professional development to support changes in instructional strategies.

[11] The study used hierarchical linear modeling (HLM) and controlled for student socio-economic status to determine the impact of improved alignment of teaching and assessment methods.

71. Not all performance-assessments provide information needed to shape instruction. In other words, a change in assessment format is not in and of itself sufficient. Several studies in the US have examined the validity of different performance-based assessments and found that often they are not aligned with contemporary research on learning and do not always test the skills and processes intended (Baxter and Glaser, 1998; Hamilton *et al.*, 1997; Pellegrino *et al.*, 1999). These studies point to weaknesses in the design of specific assessments rather than performance-based approaches, *per se*.

72. Mislevy and colleagues (1998) argue that, in order to address potential validity issues, test developers should first set out the key inferences they want to make at the beginning of the process, and then consider the different performance tasks that would provide evidence of student capabilities. Linn, Baker and Dunbar (1991) have suggested that validity criteria for performance-based assessments should include: cognitive complexity (*i.e.* the intellectual demands of tasks), content quality and coverage (*i.e.* subject matter content must be accurate and meet prevailing standards), generalisability, cost and efficiency.

73. One approach to resolving problems in balancing validity, generalisabilty and reliability has been to combine multiple choice and performance-based assessments (known as complex assessments). (As noted above, Canada and Korea both report that they use a combination of multiple choice and performance-based assessments.) Pellegrino and colleagues (1999) observe, however, that complex assessments are only a temporary fix to the challenges of designing large-scale assessments that can inform instruction.

74. Based on information provided in UNESCO and Eurydice country reports, it appears that OECD countries have not placed a strong emphasis on systematic external evaluations of large-scale, standards-based assessments. External evaluations would provide valuable information as to whether assessments are effectively aligned with standards for learning, whether assessment data are delivered in a timely manner and are being used as intended, and so on. External evaluations might also provide valuable information on the overall effectiveness of assessment and evaluation frameworks.

SECTION 5: TEACHER APPRAISAL

75. The major part of this report focuses on direct assessment of student performance – both formative and summative. This section addresses the role of teacher appraisal, as teacher performance is very directly concerned with student achievement. Indeed, several studies have shown that teacher quality is the most important school-based factor influencing student performance (Goldhaber *et al.*, 1999; Hanushek, 1992; Rivkin *et al.*, 2005; Rockoff, 2004).

76. But teacher performance appraisal appears to be relatively low priority in many OECD countries (see Annex 3). An OECD (2005) review of teacher policy found that teachers were not evaluated on a regular basis in half of the 25 countries participating in the project. According to findings of the OECD (2009a) Teaching and Learning International Survey (TALIS)[12], in most education systems, school evaluations and teacher assessments do not have a clear focus on specific aspects of education or teaching – with the exception of teachers working with students in special education and/or in multicultural settings. Almost all teachers participating in the TALIS survey agreed that school leaders do not use effective methods to assess their performance, and three-quarters of teachers reported that improvements in the quality of their teaching are not recognised (OECD, 2009).

77. In principle, strong teacher appraisal systems could serve as a powerful way to provide formative feedback to teachers – reinforcing effective teaching and assessment practices and identifying areas for improvement. Teachers responding to the OECD TALIS survey indicated that they place more emphasis on those areas of practice that are emphasised in the teacher appraisal system. The survey found a statistically significant relationship in all participating countries between emphases in teacher appraisal and feedback, and influences on teacher practice.

78. Baker and colleagues (2010) also note progress in the development of standards-based appraisals of teaching practice and structured performance assessments of teachers. The model takes a formative approach. It includes a comprehensive model of goals for what teachers should know and be able to do, includes explicit standards in multiple domains for multiple levels of performance, and has detailed behavioural rating scales. It also involves collection of evidence, such as lesson plans and samples of student work, and frequent observations of classroom practice (Milanowski *et al.*, 2004). The use of such appraisal systems in some US school districts has been linked to improvements in teacher effectiveness and student achievement gains (Baker *et al.*, 2010).

79. At the same time, effective models for in-depth evaluation specifically focused on teachers' formative assessment practices are only in the early stages of development. For example, Herman and colleagues (2010) have piloted a model for evaluation of formative assessment practice which focuses on connections between and among the learning construct(s) to be measured, the task(s) teachers design to elicit student responses, and the interpretive frameworks they use to make sense of those responses and to shape subsequent instruction. Based on findings of the first year of the study, the researchers suggest that there are a number of difficulties in developing valid measures of teacher practice. It may be important, they suggest, to differentiate between teachers' engagement in the process of formative assessment and the

[12] TALIS surveyed teachers and school heads in 16 OECD countries and 7 partner countries. Data are based on respondents' self-reports.

validity of the inferences they are able to draw from that process. They also suggest that, given the high level of skills needed to integrate formative assessment into regular practice, systems will need to invest in additional support. Effective appraisals and evaluations will help to identify those areas where teachers most need to develop their skills.

SECTION 6: STRENGTHENING THE LINKS BETWEEN LARGE-SCALE, STANDARDS-BASED ASSESSMENTS AND CLASSROOM-BASED FORMATIVE ASSESSMENT

80. Thus far, the discussion has focused on barriers to using data from large-scale, standards-based assessments to diagnose individual student learning needs and shape instruction. A number of different approaches have been proposed to address these barriers. They include efforts to:

- Strengthen teachers' assessment roles
- Strengthen teacher appraisal
- Draw on advances in cognitive sciences to strengthen the quality of both formative and summative assessment
- Develop curriculum embedded or "on-demand" assessments.
- Develop complementary diagnostic assessments for students at lower proficiency levels to identify specific learning difficulties.
- Consider administering large-scale assessments developed primarily for monitoring purposes to a sample of students rather than to every student
- Take advantage of developments in technology-based assessments

6.1 Strengthen teachers' assessment roles

81. External assessments help to ensure that schools are working toward central standards. But, as discussed above, a number of studies point to limits of validity and reliability of large-scale assessments. Some commentators have suggested that it is better to blend these external assessments with teachers' classroom-based assessments, ensuring a level of accountability but also providing room for teachers' professional judgement (Darling-Hammond and McCloskey, 2008; Janssens and van Amelsvoort, 2008).

82. Because teachers are able to observe students' progress toward the full range of goals set out in standards and curriculum over time and in a variety of contexts, their assessments may help to increase validity (Harlen, 2006). Moreover, as Shepard (2000) argues, teachers using formative assessment make quick corrections; an incorrect assessment of a student's learning on one day may be adjusted according to information gathered in subsequent interactions. Stronger assessment roles for teachers may also help to build their assessment literacy and skills, ensure closer links between assessment and instruction, and strengthen their professionalism.

83. Several OECD countries and regions already involve teachers in both the development and scoring of graduation examinations. For example Denmark, Greece, the Netherlands, Norway, Poland, Sweden, Switzerland and the UK combine students' scores on external school-leaving examinations with teachers' assessments (see Annex 1). As noted above, there are some concerns regarding reliability of performance-based assessments. For example, the Swedish Schools Inspectorate found that teachers' scoring of national assessments does not meet criteria for reliability (Nusche *et al.*, 2011), suggesting that

teachers' grading/scoring of students' classroom performance is also highly variable. However, Caldwell, Thorton and Gruys (2003) found that training helps to increase reliability of scores. New ICT programmes that are able to score open-ended performances are also under development, and may facilitate the work of human raters.

84. Participation in the development and/or scoring of assessments can also serve as an important form of professional development for teachers. More generally, it is important to build teachers' assessment literacy, and to ensure that data from external examinations are delivered to schools in a form accessible to teachers and school leaders. Assessment literacy includes awareness of the different factors that may influence the validity and reliability of the results – and capacity to make sense of data, identify appropriate actions and track progress (Earl and Fullan, 2003; Fullan, 2001). Lachat and colleagues (2006) have found that teachers increase their assessment literacy when they organise data around key questions, have access to disaggregated data, and use work in teams or with a data coach.

6.2 Strengthen Teacher Appraisal

85. While OECD countries do not currently place a strong emphasis on teacher appraisal, there are a few examples of effective approaches. These include protocols that use research-based criteria for effective practice. The protocols may be used for classroom observations or examination of videotapes of classroom practice, or for review of lesson plans and samples of student work. They may also call for review of how a teacher's instruction affects student learning over time. Appraisals of teachers' work may be performed by competent supervisors and may include peer review, as well (Baker *et al.*, 2010).

86. Protocols of teaching practice may also include measures on the effectiveness of teachers' formative assessment practice. However, there are a number of challenges to developing coherent and valid measures of formative assessment practice, as it involves several steps, including the assessment process, interpretation of evidence of student learning, and the development of next steps for instruction.

87. There is a real need for further research and pilot projects to test alternative approaches to teacher appraisal. If appraisals are to serve a formative purpose for teachers, then they should be considered as part of a coherent approach to supporting individual professional development as well as to meeting student needs as identified in broader assessment and school-level evaluations.

6.3 Draw on advances in cognitive sciences to strengthen both formative and summative assessment

88. Based on current knowledge of learning and advances in measurement theory, it is possible to develop strong summative assessments that can also shape instruction and classroom-based formative assessments.

89. Chudowsky and Pellegrino (2003) suggest that effective summative assessments should:

- Be based on empirical evidence of how students learn in a given domain. Targets of inference should include typical errors or misconceptions in the domain, which provide insights into student thinking and which might be addressed in subsequent teaching and learning.

- Focus on cognitive demands rather than specific content so that assessments are more effectively aligned with curricula that promote higher-order thinking, including problem solving and reasoning.

- Provide criteria by which to differentiate between levels of performance in the domain (from novice to highly competent), and be based upon the central concepts students must understand in order to make further progress.

- At the same time, assessments should allow for a variety of ways to value different kinds of learning performance (*e.g.* different kinds of tasks).

90. Each of these points is relevant to large-scale, standards-based assessments as well as classroom-based assessments. Indeed, well-designed standards-based assessments that focus on core concepts (and not just those that are easiest to measure), follow student reasoning processes, and include questions to identify typical misconceptions may serve as useful models for classroom-based assessments.

91. In turn, well-designed classroom-based assessments may serve useful complements to standards-based assessments because teachers have more opportunities to track different kinds of student performance and to analyse patterns that might reveal specific weaknesses or misconceptions.

6.4 Develop curriculum-embedded or "on-demand" assessments

92. Curriculum-embedded assessments may help to address several of the challenges of developing assessments that are instructionally useful. Curriculum-embedded assessments avoid problems of generalisability and reliability associated with teacher-designed assessments. Well-designed curriculum-embedded or on-demand assessments may also help improve the validity of teachers' assessments – helping to ensure that teachers are able to make appropriate inferences about student learning in relation to learning goals. They also provide information in a timely manner – essential if the results are also to be used for formative assessment.

93. Both Sweden and Scotland 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 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. Centrally developed assessments are also available. The on-demand assessment results may comprise up to 50% of final examination scores (Darling-Hammond and McCloskey, 2008; OECD, forthcoming; Sliwka and Spencer, 2005).

94. Shavelson and colleagues (2008) have developed a system of curriculum-embedded formative assessments for a popular science curriculum for lower secondary school students. The programme was field-tested in a small, randomised trial. They found that the process of embedding assessments within the curriculum helped to clarify teaching goals, as well as to identify inconsistencies within and between lessons. Their tentative conclusion was that the embedded assessments, when used as intended, could enhance student performance. They also noted that collaboration between curriculum and assessment experts as well as with teachers was sometimes challenging, but also essential for the success of the project.

95. While curriculum-embedded and on-demand assessments are currently aimed primarily at classrooms (*i.e.* the results do not feed into large-scale assessments developed for school accountability or improvement purposes) they do help to ensure a much closer alignment between assessments developed for different purposes. Potentially, these test data may also be used to for monitoring and accountability purposes.

6.5 Use diagnostic assessments for students at lower proficiency levels to better identify specific learning needs

96. The idea that data from large-scale, external assessments should be used for improvement is key to standards-based education. In some cases, as in France and Spain, where assessments are administered early in the academic year, they may also serve a diagnostic purpose. In France, the assessments are administered to students who are making key transitions in their schooling. Trends apparent in the aggregate data help to shape policy and identify areas where the majority of students are performing below expectations. Note that the use of the term "diagnostic" for these large-scale assessments is not used in its clinical sense. Rather, they help to diagnose areas of weakness across student cohorts, where a policy response may be appropriate.

97. On the other hand, while standards-based assessments may signal which students are having difficulty, they cannot identify the source of individual difficulties. As noted above, standards-based assessments are usually reported according to criterion-referenced proficiency classifications. The classifications are usually very broad, and mask significant level of heterogeneity. At the same time, it is probably not necessary to develop large-scale assessments with diagnostic capabilities. Rather, teachers may draw upon existing batteries of diagnostic assessments for those students who perform poorly on standards-based or other assessments in order to identify the source of learning difficulties and develop appropriate instructional responses.

98. In France, the Assessment, Prospects and Performance Directorate (DEPP – *Direction de l'Évaluation, de la Prospective et de la Performance*) of the Ministry of National Education has developed a number of support tools to support diagnostic assessment of individual student needs in a range of subjects and at all levels. These assessments may be administered at any point in the year. The key issue here is to recognise the limits of standards-based assessment for diagnosis of individual student needs, but to also develop a separate strategy for developing appropriate testing strategies for students who are having difficulties, and to provide remedial teaching when necessary.

6.6 Consider population sampling for large-scale assessments used for monitoring purposes

99. Most countries implementing standards-based assessments require that all students in given years participate in the assessments. However, countries may choose to administer assessments to a smaller sample of students in order to monitor trends in student performance across different areas of the curriculum. Currently, as noted in Section 3, Finland and Korea take this approach.

100. Population sampling may prevent the problem of "teaching to the test" – particularly if, as Wiliam (2001) suggests, different students are tested on different tasks. This approach would also provide more accurate information for policy purposes – both because it would be possible to assess student performance in all areas of the curriculum, and because it would help to avoid score inflation associated with "teaching to the test". Systems could still meet demands to hold schools and teachers accountable for meeting goals for learning through school inspection and teacher appraisals. Systems might also administer assessments to samples of randomly selected students in every school. The latter approach would focus teachers' attention on central standards, but lower incentives to teach to the test. It would also provide sufficient data for systems to identify schools in need of additional support, and provide data needed for school level decisions related to resource allocation, areas where groups of students may need additional support, and so on.

101. However, to the extent that systems are able to improve integration of large-scale, standards-based assessments and classroom-based formative assessment, the cost of administering assessments to every student in a given cohort will be justified. Tests that align with goals for higher order learning,

provide more detailed data and deliver results in a timely manner will provide important data to guide classroom instruction as well as policy.

6.7 Take advantage of technology

102. Chudowsky and Pellegrino (2003) suggest that new technologies may facilitate classroom-based formative assessment while also generating information for external assessment needs. It would be possible, they argue, to gather data on how students approach a range of learning activities in electronic formats. These data would be useful for classroom-based formative assessment, and potentially could feed into summative assessments. Summative assessments based on these multiple observations would provide more accurate information on student performance, as the data would be gathered over time, for a range of tasks and in different contexts. However, as Chudowsky and Pellegrino caution, there are a number of concerns regarding privacy, equity and practicality that would need to be addressed before such a system can be implemented. Technology nevertheless provides a number of possibilities in regard to gathering data in a more formative fashion.

103. In addition, 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). For example, new ICT-based assessments may incorporate simulation activities, or allow students to interact and collaborate on constructed response formats. With some assessments, students may receive feedback on their work while they are taking the test (Lewis, 1998).

104. These ICT-based programmes offer the possibility for efficient scoring of large-scale, performance-based assessments that are more effectively aligned with higher order skills emphasised in countries' national standards and curricula, with timely feedback. 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).

105. These kinds of assessments are relatively new, and as of yet, relatively limited in number. 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 sure to accelerate.

SECTION 7: GENERAL POLICY IMPLICATIONS AND CONCLUSIONS

106. This report has surveyed a broad range of issues related to assessment and evaluation in OECD countries. It has uncovered a number of gaps in research and pointed to some proposed remedies. This final section sets out the broad implications for policy and research.

7.1 Learn from the bottom up: use formative assessment data to build knowledge about what works in policy and practice

107. If systems are to adapt and improve, feedback on the processes and factors associated with successful student outcomes is vital. Indeed, one of the strong critiques of top-down approaches to external summative assessment is that while they set out incentives to improve student outcomes, teachers and school leaders receive little guidance on how to do so.

108. The iterative nature of formative assessment provides opportunities to develop more nuanced views about how students learn and adapt. Moreover, data gathered in formative processes provide more detail and in a more timely fashion. These data are more useful for testing hypotheses about the impact of different practices (O'Day, 2002), and how these practices might fare in different environments. While data gathered through external summative assessments might point to worthwhile areas for further investigation, they are not always sufficient to permit meaningful interpretations of what works, for whom, and under what circumstances.

7.2 Promote teacher professionalism

109. Formative assessment has been shown to be a highly effective approach to improving student achievement, but it also requires a high level of skill on the part of teachers. Teachers need to develop skills not only to identify individual student learning needs, but also to respond to them. Both subject matter and pedagogical knowledge come into play.

110. Systems need to make significant and sustained investments in teacher professional development to support effective practices such as formative assessment. Professional development should be targeted clearly to areas of need as identified in appraisals as well as school-level evaluations (both internal and external). But real change will require opportunities for teachers to learn, reflect and experiment with new approaches, and to work with colleagues.

111. Strong teacher appraisal systems, linked to priorities for system and school improvement, can also play a significant role in promoting teacher professionalism. At this point, there are few models of strong teacher appraisal systems in OECD countries. But the growing body of research on the qualities of effective teachers and teaching methods provide a sound foundation for development.

7.3 Ensure cost effectiveness by developing more effective approaches to assessment

112. Policy makers need to consider investments in strengthening the quality of assessment (both large-scale, standards-based assessments and classroom-based formative assessments). The majority of the proposals for strengthening assessment made in this report are reasonably cost effective and address many of the problems with existing assessment systems.

113. Among the most costly of these proposals are the use of human raters to score performance-based assessments, and new ICT-based assessments to measure complex performances. But there are also significant pay-offs associated with these approaches. Rating panels provide teachers with valuable professional development experience. There is also evidence that the validity, reliability and generalisability of assessment scores are quite high when human raters are well trained. And, more sophisticated ICT-based assessments, particularly those that may be administered at different points throughout the year, provide significant opportunities to improve the integration of formative and summative assessment and to improve learning.

7.4 Address Gaps in Research and Development

114. This paper has revealed significant advances in research on learning, formative and summative assessments, as well as a number of gaps. Priorities for further research and development include:

- Understanding what kinds of policies, training programmes and other supports will help bring effective classroom-based formative assessment to scale.

- Deepening understanding of how students learn in different subject domains, and implications for teaching and assessment.

- Developing educational measurements that reflect current conceptions of learning and the development of competence, while also meeting criteria for validity, reliability and generalisability.

- Developing and piloting new teacher appraisal systems. Appraisals should address the knowledge and skills needed for effective formative assessment practice, and the ability to adjust teaching in different subject domains to meet student needs.

- Exploring ways in which technology can support and promote both classroom-based assessment (formative and summative) as well as large-scale, standards-based assessments. Computer-based technologies offer ways for students to demonstrate problem-solving processes, and provide timely feedback on learning progress.

- Encouraging collaboration – among policy makers, teachers, practitioners, as well as researchers working in domains of curriculum development and assessment.

REFERENCES

Abelmann, C.H., R.F. Elmore, J. Even, S. Kenyon and J. Marshall (1999), *When Accountability Knocks, Will Anyone Answer?*, University of Pennsylvania, Consortium for Policy Research in Education, Philadelphia.

Abrams, L.M., J.J. Pedulla and G.F. Madaus (2003), "Views from the Classroom: Teachers' Opinions of Statewide Testing Programs", *Theory into Practice*, Vol. 42, pp. 19- 29.

Allal, L. (1979), "Stratégies d'évaluation formative : conceptions psycho- pédagogiques et modalités d'application", in L. Allal, J. Cardinet and P. Perrenoud (eds.), *L'évaluation formative dans un enseignement différencié*, Peter Lang, Bern, pp. 153-183.

Allal, L. and G. Schwartz (1996), "Quelle place pour l'évaluation formative dans l'enseignement au cycle d'orientation ?", *CO Infos,* No. 178, pp. 5-8.

Allal, L. and L. Mottier-Lopez (2006), "Formative Assessment of Learning: A Review of Publications in French", in OECD, *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.

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.

Audibert, S. (1980), "En d'autres mots...l'évaluation des apprentissages !", *Mesure et évaluation en éducation,* Vol. 3, pp. 59-64.

Baker, E. *et al.* (2010), "Problems with the Use of Student Test Scores to Evaluate Teachers", *EPI Briefing Paper No. 278*, Economic Policy Institute, Washington, DC.

Baxter, G.P. and R. Glaser (1998), "The Cognitive Complexity of Science Performance Assessment in the Science Classroom", *Educational Measurement: Issues and Practice*, Vol. 17, No. 3, pp. 37-45.

Bell, B. and B. Cowie (2001), *Formative Assessment and Science Education*, Kluwer Academic Press, Dordrecht, The Netherlands.

Black, P. (1993), "Formative and Summative Assessments by Teachers", *Studies in Science Education*, Vol. 21, pp. 49-97

Black, P. (2002), "Research and Development of Educational Assessment*", Oxford Review of Education*, Vol. 26, pp. 407-419.

Black, P. and D. Wiliam (1998), "Assessment and Classroom Learning", *Assessment in Education: Principles, Policy and Practice,* CARFAX, Oxfordshire, Vol. 5, pp. 7-74.

Black, P. and D. Wiliam (2003), "In Praise of Educational Research: Formative Assessment", *British Educational Research Journal*, Vol. 29, pp. 623-637.

Black, P., C. Harrison, C. Lee, B. Marshall and D. Wiliam (2002), *Working Inside the Black Box: Assessment for Learning in the Classroom*, Department of Education and Professional Studies, King's College, London.

Bloom, B.S. (1968), "Learning for Mastery", *Evaluation Comment,* Vol. 1, pp. 1-12.

Bloom, B.S., J.T. Hasting and G.F. Madaus (1971), *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw- Hill Book Co, New York.

Boulet, M., G. Simard and D. Demelo (1990), "Formative Evaluation Effects on Learning Music", *Journal of Educational Research*, Vol. 84, pp 119-125.

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

Brown, J. and P. Duguid (2000), "Balancing Act: How to Capture Knowledge without Killing It", *Harvard Business Review*, Vol. 78, pp. 73-80.

Buly, M.R. and S.W. Valencia (2002), "Below the Bar: Profiles of Students Who Fail State Assessments", *Educational Evaluation and Policy Analysis*, Vol. 24, pp. 219-239.

Butler, R, (1998), "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.

Caldwell, C., C.G. Thorton and L.M. Gruys (2003), "Ten Classic Assessment Center Errors: Challenges to Selection Validity", *Public Personnel Management*, Vol. 32, pp. 73-88.

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.

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, CT, pp. 7-40.

Crooks, T.J. (1988), "The Impact of Classroom Evaluation on Students", *Review of Educational Research*, Vol. 58, pp. 438-481.

Darling-Hammond, L. and L. McCloskey (2008), "Assessment for Learning around the World: What Would It Mean to Be Internationally Competitive?", *The Phi Delta Kappan*, Vol. 90, pp. 263-272.

Duschl, R.D. and D.H. Gitomer (1997), "Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms, *Educational Assessment*, Vol. 4, pp. 37-73.

Earl, L. and M. Fullan (2003), "Using Data in Leadership for Learning", *Cambridge Journal of Education*, Vol. 33, pp. 383-394.

Fullan, M. (2001), *Leading in a Culture of Change*, Jossey-Bass, San Francisco.

Gauld, C.F. (1980) "Subject Oriented Test Construction", *Research in Science Education*, Vol. 10, pp. 77-82.

Gipps, C. (1996), "Assessment for Learning", in A. Little and A. Wolf (eds.), *Assessment in Transition*, Pergamon, Elsevier Science Ltd., Oxford, UK.

Goldberg, G.L. and B.S. Roswell (2000), "From Perception to Practice: The Impact of Teachers' Scoring Experience on Performance-based Instruction and Classroom Assessment", *Educational Assessment*, Vol. 6, pp. 257-290.

Goldhaber, D., D. Brewer and D. Anderson (1999), "A Three-Way Error Components Analysis of Educational Productivity", *Education Economics,* Vol. 7, pp. 199-208.

Griffin, P. (2007), "The Comfort of Competence and the Uncertainty of Assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 87-99.

Grubb, W.N. (2000), "Opening Classrooms and Improving Teaching: Lessons from School Inspections in England", *Teachers College Record*, pp. 696-723.

Hamilton, L.S., E.M. Nussbaum and R.E. Snow (1997), "Interview Procedures for Validating Science Assessments", *Applied Measurement in Education*, Vol. 10, pp. 181-200.

Hanushek, E. (1992), "The Trade-off Between Child Quantity and Quality", *Journal of Political Economy,* Vol. 100, pp. 84-117.

Hanushek, E.A. and M.E. Raymond (2005), "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, Vol. 24, pp. 297-327.

Harlen, W. (2006), "Criteria for Evaluating Systems for Student Assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 15-28.

Harlen, W. and M. James (1997), "Assessment and Learning: Differences and Relationships between Formative and Summative Assessment", *Assessment in Education*, Vol. 4, pp. 365-379.

Heritage, M. (2010), Formative Assessment: Making It Happen in the Classroom, Corwin Press, Thousand Oaks, CA.

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.

Honig, B. (2001), Comment on Chester E. Finn, Jr., "State Academic Standards" *Brookings Papers on Education Policy*, No. 4, pp. 174-179.

Hopkins, D. and D. Reynolds (2001), "The Past, Present and Future of School Improvement: Towards the Third Age", *British Educational Research Journal*, Vol. 27, pp. 459-475.

Jacob, B.A. (2003), "Getting Inside Accountability: Lessons from Chicago", *Brookings-Wharton Papers on Urban Affairs*, pp. 41-81.

Janssens, F. and G. van Amelsvoort (2008), "School Self-evaluations and School Inspections in Europe: An Exploratory Study", *Studies in Educational Evaluation,* Vol. 34, pp. 15-23.

Janssens, F.J.G. (2005), "School Inspections and Educational Governance", Inaugural Lecture, University of Twente, Enschede (cited in Janssens and van Amelsvoort, 2008).

Janssens, F.J.G. (2007), "Supervising the Quality of Education", in W. Bottcher and H.G. Kotthoff (eds.), *Schulinspektion: Evaluation, Rechenschaftsleging und Qualitatsentwicklung*, Waxman, Münster (cited in Janssens and van Amelsvoort, 2008).

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.

Köller, O. (2005), "Formative Assessment in Classrooms: A Review of the Empirical German Literature", in OECD, *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.

Koretz, D., B., Stecher, S. Klein and D. McCaffrey (1994), "The Perceived Effects of the Maryland School Performance Assessment Program", *CSE Technical Report 409*, Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

Lachat, M.A., M. Williams and S.C. Smith (2006) "Making Sense of ALL Your Data", *Principal Leadership*, October, pp. 16-21.

Linn, R.L. (2000), "Assessments and Accountability", *Educational Researcher,* Vol. 29, pp. 4-16.

Linn, R.L., E.L. Baker and S.B. Dunbar (1991), "Complex, Performance-based Assessment: Expectations and Validation Criteria", *Educational Researcher*, Vol. 20, pp. 5-21.

Looney, J. (2009), "Assessment and Innovation in Education", *OECD Education Working Papers No. 24*, OECD Publishing, Paris, doi: 10.1787/222814543073

MacBeath, J., D. Meuret, M. Schratz and L.B. Jakobssen (1999), *Evaluating Quality in School Education: A European Pilot Project*, Final Report, European Commission Education and Training Youth, Office for Official Publications of the European Communities, Luxembourg.

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 (pp. 83-121). Chicago: University of Chicago Press.

Matthews, P. and P. Sammons (2004), *Improvement through Inspection: An Evaluation of the Impact of Ofsted's Work*, Ofsted/University of London, London.

McDonnell, L.M. and C. Choisser (1997), "Testing and Teaching: Local Implementation of New State Assessments", *CSE Technical Report 442*), Center for Research on Evaluation, Standards, and Student Testing, Los Angeles.

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.

Mischo, C. and F. Rheinberg (1995), "Erziehungsziele von Lehrern und individuelle Bezugsnormender Leistungsbewertung", *Zeitschrift für Pädagogische Psychologie*, Vol. 9, pp. 139-151.

Mislevy, R.J. (1993), "Foundations of a New Test Theory", in N. Fredericksen, R.U. Mislevy and I.I. Bejar (eds.), *Test Theory for a New Generation of Tests*, Erlbaum, Hillsdale, NJ, pp. 19-39.

Mislevy, R.J., L.S. Steinberg, F.J. Breyer, R.G. Almond and L. Johnson (1998), *A Cognitive Task Analysis, with Implications for Designing A Simulation- Based Performance Assessment*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

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.

New Zealand Ministry of Education (2010), "Position Paper: Assessment, Schooling Sector", September 24, *www.minedu.govt.nz/~/media/MinEdu/Files/TheMinistry/AssessmentPositionPaperSep2010.pdf*

Newmann, F.M., M.B. King and M. Rigdon (1997) "Accountability and School Performance: Implications from Restructuring Schools", *Harvard Educational Review*, Vol. 67, pp. 41-74.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD, Paris, available at *www.oecd.org/dataoecd/38/42/47169533.pdf.*

O'Day, J. (2002), "Complexity, Accountability and Improvement", *Harvard Educational Review*, Vol. 72, *http://gseweb.harvard.edu/~hepg/oday.html*.

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000*, OECD, Paris.

OECD (2005), *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.

OECD (2009a), *Creating Effective Teaching and Learning Environments: First Results from the Teaching and Learning International Survey (TALIS)*, OECD, Paris.

OECD (2009b), *Education at a Glance: OECD Indicators*, OECD, Paris.

Pellegrino, J.W., G.P. Baxter and R. Glaser (1999), "Addressing the 'Two Disciplines' Problem: Linking Theories of Cognition and Learning with Assessment and Instructional Practice", *Review of Research in Education*, Vol. 24, pp. 307-353.

Perrenoud, P. (1998), "From Evaluation to a Controlled Regulation of Learning Processes: Towards a Wider Conceptual Field", *Assessment in Education: Principles, Policy and Practice,* CARFAX, Oxfordshire, Vol. 5, pp. 85-102.

Queensland Studies Authority (QSA) (2001), "Externally Moderated School-based Assessment: Myths and Facts", QSA, Australia, *www.qsa.qld.edu.au/downloads/approach/ext_mod_school-based_myths.pdf*, accessed 2 April 2010.

Rivkin, S., E. Hanushek and J. Kain (2005), "Teachers, Schools, and Academic Achievement", *Econometrica,* Vol. 73, pp. 417-458.

Rockoff, J. (2004), "The Impact of Individual Teachers on Students' Achievement: Evidence from Panel Data", *American Economic Review*, Vol. 94, pp. 247-252.

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.

Sadler, R. (1989), "Formative Assessment and the Design of Instructional Systems", *Instructional Science*, Vol. 18, pp. 119-144.

Scriven, M. (1967), "The Methodology of Evaluation", *AERA Monograph Series on Evaluation*, Vol. 1, pp. 39-83.

Shavelson, R., *et al.* (2008), "On the Impact of Curriculum-embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers", *Applied Measurement in Education*, Vol. 21, pp. 295-314.

Shepard, L.A. (2000), "The Role of Assessment in a Learning Culture", *Educational Researcher*, Vol. 29, pp. 4-14.

Shepard, L.A. (2006), "Classroom Assessment", in R.L. Brennan (ed.), *Educational Measurement* (4th ed.), American Council on Education/Praege Publishers, Westport, CT, 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.

Shepard, L.A., K. Hammerness, L. Darling-Hammond and F. Rust (2005), "Assessment" , in L. Darling-Hammond and J. Bransford (eds.), *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do,* Jossey-Bass, San Francisco, pp. 275-326.

Sliwka, A. and E. Spencer (2005), "Scotland: Developing a Coherent System of Assessment", in OECD, *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.

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.

UNESCO, World Data on Education Database, International Bureau of Education, *www.ibe.unesco.org/en/services/resources-studies/world-data-on-education.html*, Reports accessed Spring 2010.

Vanhoof, J. and P. Van Petegem (2007), "Matching Internal and External Evaluation in an Era of Accountability and Development: Lessons from a Flemish Perspective", *Studies in Educational Evaluation*, Vol. 33, pp. 101-119.

Wiliam, D. (2001), "What is Wrong with Our Educational Assessments and What Can Be Done About It?", *Education Review*, Vol. 15, pp. 57-62.

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.

Williams, J. and J. Ryan (2000), "National Testing and the Improvement of Classroom Teaching: Can the Coexist?", *British Educational Research Journal*, Vol. 26, pp. 49-73.

Yoon, B. and L.B. Resnick (1998), "Instructional Validity, Opportunity to Learn and Equity: New Standards for Examinations for the California Mathematics Renaissance", *CSE Technical Report No. 484*, Center for Research on Evaluation, Standards and Student Testing, University of California, Los Angeles.

ANNEX 1: ASSESSMENT AND EVALUATION FRAMEWORKS OECD COUNTRY POLICIES

The information contained in this annex was drawn from the UNESCO International Bureau of Education's "World Data on Education" (WDE) reports for 2006 and Eurydice's country profile reports. The full WDE reports are available at http://www.ibe.unesco.org/Countries/WDE/2006/index.html and Eurydice's Eurybase country profile reports are available at http://eacea.ec.europa.eu/education/eurydice/eurybase_en.php.

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
AUSTRALIA	States and territories assess students using their own state-based test instruments measuring student performance against nationally agreed benchmarks in numeracy and literacy for grades 3, 5 and 7. The results are reported in the Annual national report on schooling (ANR). States and territories have their own external accreditation process for certifying school completion at Year 12, and ranking students for entry into year tertiary institutions. Several states also provide certification at Year 10, using a state-based or school-moderated assessment process.	States and territories use different school self-evaluation tools and rating scales.	No external inspection agency. Federal funding for schools is linked to national performance measures and targets. (UNESCO, 2006)
AUSTRIA	Students in academic upper secondary schools, higher-level secondary technical and vocational education colleges take matriculation and diploma examinations. There are no formal external tests during compulsory school.	Schools are required to develop 'school programmes' outlining objectives and action plans. The federal Minister of Education initiated the Quality in Schools (QIS) programme in 1996. The aim of the programme is to encourage and support schools in critical appraisal, monitoring and development. It defines specific aims and measures to be taken in different areas.	School inspections are directed at regional level. There is no central school inspectorate. (UNESCO, 2007)
BELGIUM (FL. CMMTY)	The Flemish Community does not have any centrally administered examinations. Inter-school tests are organized each year (municipal or inter-diocesan) for certain groups of subjects, such as the mother language and arithmetic in the final year of primary school. Test results may not be published. The tests are voluntary. By law, test results may not be published on a comparative basis. The Inspectorate may organize tests in certain learning areas.	Schools are required to develop school plans defining basic educational choices and concrete actions. The *pouvoir organisateur* produces annual activity reports on the previous year.	The Inspectorate makes a comprehensive analysis of each school at each level, assessing attainment of general objectives. There are no systematic evaluations of individual schools. (UNESCO, 2007).

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
BELGIUM (FR. COMMTY)	In 2002, the Community introduced an initiative to set up standardised external assessment for students at the beginning of each stage of education. The *pouvoir organisateur* at each level of education sets out criteria for assessment of schoolwork.		(Eurydice, 2008/09)
CANADA	The Ministry or Department of Education in some provinces (*e.g.* Québec, Alberta, British Columbia) conduct province-wide school-leaving examinations. Other provinces (*e.g.* Newfoundland and Labrador) conduct standardized tests, such as the Canadian Test of Basic Skills) to monitor student performance. In 1989, the Council of Ministers of Education, Canada (CMEC) introduced the first national student assessment, the School Achievement Indicators Program (SAIP) to track student achievement in primary and secondary reading, writing and mathematics. The SAIP is administered to a random sample of 13- and 16-year-old students in all provinces and territories (except Saskatchewan). Several provinces have also introduced large-scale accountability testing to measure systems, schools and/or individual students.		Inspection of schools and school districts varies from province to province. (UNESCO, 2006)
CZECH REPUBLIC	Students take a final examination leading to award of apprentice certificate or diploma, or maturita exam, leading to award of a general certificate of upper secondary education. Diversity in the content and quality of maturita examinations led to reforms. A new model combines a common test for all students, and an internal, specialized test. Results of the common test are used to compare outcomes of general secondary education. There are no national performance tests.	Since 2006/07, all schools have been required to conduct self-evaluations	The Czech School Inspectorate (CSI) evaluates educational process and outcomes, as well as human resources, facilities, and financial resources. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
DENMARK	Students take (non-compulsory) matriculation examinations at the end of Years 9 and 10. There are standard rules for examinations to ensure uniformity throughout the country. There is an examination committee for each subject, consisting of teachers and the Department's subject advisors. Examinations are held in subjects selected (at random) by the Department for each individual school. Students receive a mark for the year's work in the subject area, as well as for examination results The average of the two sets of marks is then the student's examination result. The government introduced national tests in 2006/07 for ten subjects, including: Danish, English, mathematics and the natural sciences. The Danish government intends to introduce IT-based examinations.	No official school self-evaluation policy noted. All educational institutions are required by law to have a website providing information about the pedagogical principles guiding the instruction, grade averages for individual subjects and levels, and any relevant other information on the quality of the instruction.	The Evaluation Institute (EVA) is an independent institution under the Ministry of Education. EVA develops methods for evaluation of the teaching and learning processes, and advises and collaborates on issues related to quality. (UNESCO, 2007)
FINLAND	Matriculation examinations are set and assessed by a national committee appointed by the Ministry of Education. The examinations include four compulsory subjects, of which the mother tongue (either Finnish or Swedish, depending on the language of instruction at the school) is obligatory. Other subject may be the second official language, foreign languages, mathematics or general studies. Students also may take one or more optional examinations. The Finnish Ministry of Education and Culture confirms an evaluation plan for 5 years. The Finnish National Board of Education implements assessments of learning outcomes in a representative sample of schools. The Evaluation Council is responsible for implementing other evaluations. The results of all assessments and evaluations are published. The information is used to develop education.	Educational institutions are legally obligated to conduct self-evaluations.	National monitoring and evaluation focuses on the extent to which schools are meeting objectives set in statutes, education policy, and in the core curricula. There is no separate school inspection department. The Finnish National Board of Education is responsible for conducting national assessments concerning learning outcomes and The Evaluation Council is responsible for other evaluations projects. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
FRANCE	France introduced national standardised assessments in 1989. They are administered at the beginning of the school year, and are used for diagnostic purposes. The national Ministry of Education analyses a representative sample of the results to develop an understanding of student achievement across the country. At the age of 15, students may take the *brevet des colleges*, or for those in vocational education, the *certificat d'atitude professionelle* (CAP) and the *brevet d'études professionelles* (BEP). The main requirement for entry to higher education is the *baccalauréat* or academic, technical and vocational education. Students wishing to enter the *grands écoles* may take competitive entrance examination. The success rate of students on the *baccalauréat* is published in newspapers.	Schools conduct self-evaluation. School councils evaluate the "school project" for implementation of national pedagogical objectives. Secondary schools may also look to national indicators for guidance Centrally developed computer software includes a set of standard indicators for secondary schools. Schools may judge their own performance (on examinations, resources, school management and environment) against national levels. Schools are also encouraged to develop their own specific indicators based on local characteristics and needs.	The Ministry of Education's Office for Assessment and Forecasting evaluates costs, financing, organization, assessment of student achievement, school effectiveness, classroom practices, innovation project, and so on. The *Haut Conseil de l'Education* covers issues of assessment of acquired knowledge of pupils as well as the assessment of performance of educational institutions, or of educational practices. (Eurydice, 2008/09)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
GERMANY	The German Länder are cooperating in the development of external examinations that will deliver comparable information on student performance. Students take the *Abitur* (school-leaving) examination.	No official school self-evaluation policy noted	School inspections are provided for under the Basic Laws and the constitutions of the Länder (UNESCO, 2007)
GREECE	Students complete a project in their final year. The project is assessed by the supervising teacher and then by a 3-member teacher panel. The panel awards the final mark. The school-leaving certificate includes grades in six general "stream subjects" – including school-based assessments and scores from national examinations in the six subjects.	Teachers' Councils are responsible for developing the school self-evaluation (Eurydice)	The Evaluation Department of the Pedagogical Institute (PI), develops student assessment and evaluation activities and coordinates in-service teacher training activities. Regional directorates monitor schools in their region, and supervise the activities of teaching and administrative staff. (UNESCO, 2007)
HUNGARY	Students take school leaving examinations (Year 12 or Year 13). The examination is held in the school, and administered by an examining board including teachers from the school and chaired by a delegate from the educational authority. It includes written and oral questions. Schools may supplement the national examination with local examinations. Teachers from the school mark tests and essays. Since 1986, several Hungarian examination centres have developed surveys of student performance on standardized tests. Data gathered over a period of years allow analysis of changes and trends. Tests have been changed from a focus on acquisition of knowledge to assessment of general competencies.	Hungary's 'Comenius 2000' builds on three models. The first model involves co-operative work and school self-evaluation. The second model involves adaptation of local-institutions and the improvement of structural-development capacity. The third model is concerned with broader dissemination of ideas.	The Hungarian Ministry of Education established the National Public Evaluation and Examination Centre in 1999. The Centre's activities include organization of state examinations, contributions to regional development of education, and national control of educational measurement, evaluation and quality assurance. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
ICELAND	Iceland sponsors nationally coordinated examinations in Icelandic and mathematics in Years 4 and 7, and in Icelandic, mathematics, English and Danish in Year 10, at the end of compulsory school. The exams are marked centrally by a group of teachers selected by the Educational Testing Institute. The Ministry is also responsible developing analysis of the results of the nationally coordinated examinations to the schools. Results of school performance are published.	Schools are obligated to evaluate their performance. The evaluation is to cover teaching, administration, internal and external relationships. The school's evaluation methods are reviewed by an external party every 5 years.	External evaluations of compulsory and upper secondary schools focus on school management, student academic performance, teaching methods and their influence on achievement, communications within the school and between the school and parents. (UNESCO, 2007)
IRELAND	The National Assessment of Mathematics Achievement is administered in primary schools by the Department of Education and Science Inspectorate, with the support of the Educational Research Centre. Five NAMA assessments have been conducted since 1977. NAMA aims to identify factors associated with achievement, and identify student performance trends. Since 2005, primary schools have been required to administer standardized literacy and numeracy tests at 2 points during the primary school cycle. Students participate in an external assessment at the end of secondary school. The State Examinations Commission is responsible for development, assessment accreditation and certification.	No formal system of school self-evaluation is noted,	The Inspectorate in the Department of Education and Science (DES) evaluates educational processes and outcomes, and provides advice on education policy. Regional Subdivisions of the Inspectorate are responsible for implementation of inspection and evaluation services and related advisory activities in the five regional business units across the country. (UNESCO, 2007)
ITALY	Students sit the national school-leaving exam (maturita) at the end of upper secondary school. The exam includes 2 written tests and an oral test. It is held before a Board of Examiners appointed by the Ministry of Education. The National Institute for the Evaluation of the Education System (INVALSI) organizes assessment of student achievement at the end of Year 1 and Year 3 in primary school, and at the beginning of lower secondary school.		Inspectors work at both national and regional levels. The National Institute for the Evaluation of the Education System (INVALSI) leads the evaluation of the education system. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
JAPAN	There are no external examinations in Japan. Promotion between years and certification of completion are based on internal assessments. National academic achievement tests are being planned for students in Years 5 and 6 in 4 subjects (Japanese, social studies, mathematics and science), and for students in lower secondary school (Japanese, social studies, mathematics, science, and foreign language). Competitive university entrance examinations have a strong influence on teaching and learning in upper secondary schools. Assessments submitted by upper secondary schools are also taken into account in the University admissions process.	No official school self-evaluation is noted.	No external school evaluation noted. (UNESCO, 2006)
KOREA	Korea monitors the quality of the education system through the National Assessment of Scholastic Achievement (NASA) NASA measures student performance against the objectives outlined in the school curriculum. NASA samples 1% of all participants based on school level and region. NASA assesses higher order thinking skills. using both multiple-choice and free-response formats. It includes English listening comprehension, performance assessment in science . Results of individual school or student performances are not made public. However, the National Board of Educational Evaluation (NBEE) publishes an analysis of overall results each year.		There is an Inspector's office in the Ministry of Education (UNESCO, 2006)
LUXEMBOURG	There are no formal national-level examinations during compulsory schooling Students take a school leaving examination at the end of their schooling.	No information on school-self evaluation available in reports	School inspectors visit primary schools and report to the Ministry. There is no external inspection of secondary schools. (Eurydice, 2008)
MEXICO	Information only available in Spanish		

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
NETHERLANDS	School-leaving examinations generally include both internal and an external, national components. Schools develop internal exams, although the Ministry of Education sets the overall examination syllabus. This includes topics to be covered in each subject, how material is to be divided between the internal/school exam and the national exam. Schools may decide when to test students. These exams generally include two or more papers in each subject, as well as written, oral or practical tests (with both single-answer and open-ended questions). Internal exams are marked by the school's own staff. Students are informed of the results for those subjects in which they are also sitting national exams before the national exams begin. The school exam in HAVO (senior general secondary) and VWO (pre-university) also includes a long-term practical assignment involving two or more of the specialized subjects studied. Teachers calculate the final grades for each student in each subject by taking the average of the marks obtained in the internal or school exam and the national exam.	Schools are legally obligated to develop action plans at least every two years. Schools are expected to evaluate the quality of teaching, and to use findings as the basis for the school plan. The plan outlines teaching methods and how the school will improve teaching over the next four years. The plans also describe methods for student assessment and for reports. Plans are reviewed by the Education Inspectorate.	The Inspectorate includes a central office, and 12 regional offices. It has a statutory duty to promote the quality of education, and is expected to give institutions pointers as to how they can improve on the basis of their own quality assurance systems. The Inspectorate develops School Report Cards on every secondary school in the country, and makes them available to the public. The Report Card includes examination results, and general information about each school. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
NORWAY	The Directorate of Education is responsible for the National Quality Assessment System (NQAS) for primary and secondary education. The NQAS tracks learning results and student progress. National tests and school leaving examinations are part of the NQAS. National tests assessing pupils' basic skills in reading, writing, English, and mathematics (Years 4 and 7) were introduced in 2003/04. Results from the national tests are intended to help teachers adapt teaching methods and contents for individual students. School leaving tests include a centrally set written examination in one of three subjects: Norwegian, mathematics or English. Most pupils also take a locally organized oral examination. Teachers' marks and have the same status as examination marks and included on the student's certificate with end-of-year examination marks.	No official system for school-self evaluation noted. The School Development Programme (2005-09) initiated practical development projects in schools to promote school improvement, with external supports from the community, researchers, professional networks, and international networks.	There is no official inspectorate. (UNESCO, 2007)
NEW ZEALAND	Students earn a National Certificate of Educational Achievement (NCEA) when they have been successfully accumulated enough credits by being successfully assessed against the National Qualification Framework Standards. There are 3 levels: Level 1 (Year 11); Level 2 (Year 12), and Level 3 (Year 13). The National Qualification Standards moves away from a comparative measurement model (e.g. norm-referenced assessment) to a performance model (e.g. standards-based, or criterion-referenced assessment) National Monitoring of Standards is an external assessment. Representative samples of New Zealand students are assessed at successive points (every four years) to track trends in educational performance. Approximately 3% of 8- and 12-year old students are assessed.	No official school self-evaluations noted.	The Education Review Office (ERO) reports on the quality of education in all schools. It provides regular, independent evaluative reports for the Minister of Education, school governing authorities and managers; parents and the wider public. (UNESCO, 2006)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
POLAND	Following reforms introduced in 1999 (with implementation over a period of five years, to the 2004-05 school year), students at the end of primary school now take an external standardized test of skills and knowledge in the humanities, mathematics and natural science (as a condition for graduation from primary school. In the third year of gymnasium, students take a test in reading, writing, reasoning, use of information and practical application of knowledge. The results of the tests are not used for selection purposes. Poland introduced a new matura examination in 2004-05. It is an external standardized test, and gives access to higher education. The test includes a written component (prepared and assessed by an external examination commission) and an internal component (prepared and assessed by teachers in the school).	School self-evaluation is focused on identifying teachers' professional development needs, and improvement of school efficiency and effectiveness. The Ministry of National Education takes these evaluations into account.	The Ministry of National Education, local Kurators and school head teachers and directors provide pedagogical supervision. Parent, student and teacher views are taken into account. (UNESCO, 2007)
PORTUGAL	There are national examinations in Years 4, 6 and 9 on Portuguese, and mathematics. The tests in Years 4 and 5 are used to monitor and evaluate performance of the education system. Examinations at Year 9 are for student assessment and certification. In upper secondary school, students enrolled in scientific-humanistic courses must pass a final national examination.	No official requirements for school self-evaluation reported.	Inspection is the responsibility of the General Inspectorate of Education, which has regional delegations supervising all aspects of non-higher education. (Eurydice, 2006/07)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
SLOVAKIA	Regional school offices appoint chairs of examination boards for matura /graduation examinations and state language examinations. The matura exam consists of written and oral examinations in language and literature in the mother tongue, an oral examination in mathematics or foreign language, as well as two oral exams in optional subjects chosen by students. Students are admitted to secondary education upon completion of basic education and after having successfully passed the entrance examinations. Examinations have oral and/or written form, and usually cover Slovak language, mathematics, and sometimes foreign language. Students may apply to more than one school.		The State School Inspectorate supervises pedagogy. The Inspectorate submits reports to the Minister of Education on school activities, and proposes changes in the school network on the basis of identified shortcomings, and takes corrective measures. The Slovak Republic collaborated with the Ministry of Education in France on a September 2002 assessment of the knowledge and skills of pupils in Slovak language and mathematics. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
SPAIN	Educational institutions implement diagnostic assessment of student achievement of competencies at the completion of the fourth year of primary education, and the second year of compulsory secondary education. The aim is to provide formative guidance for institutions as well as information to families and communities. The assessments differ in content and methodologies across the Autonomous Communities, and from each year to the next. Students who have achieved basic competences are awarded the Secondary school certificate have access to the Bachillerato and intermediate vocational training. At the upper secondary level, students who have achieved acquired competencies are awarded a graduation certificate and have access to higher education. There are no large-scale external assessments of student performance	School self-evaluation has been mandatory since 1990. Schools write annual reports at the end of each school year with an evaluation of school activities and results, and develop plans for improvement. Teachers assess their own practices and processes along with student assessments. The Teachers' Assembly evaluates the educational process and the evolution of the school's academic performance, based on student assessment results. The Assembly also assesses teaching activities included in the curricular project. The education Authorities of Autonomous Communities support and facilitate school self-evaluation with awards, specific plans and assistance.	The 2006 *Ley Orgánica de Educación established the* framework for general evaluation of the education system as well as of educational institutions, and established the Institute of Evaluation (IE). Autonomous Communities are responsible for presenting their own external evaluation processes within the national plan. The model varies. It is to complement school self-evaluation Relevant bodies in Spain's Autonomous communities must collaborate in the implementation of general diagnostic assessments to gather representative data on students and schools. They also carry out specific school evaluations. The IE has been tasked with updating and annual revision of the State System of Educational Indicators. (Eurydice, 2009/10)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
SWEDEN	Upper secondary school students receive marks. Criteria for awarding marks are specified in course syllabi, which outline goals to be achieved, and minimum achievement levels. There are national tests in Swedish, English and mathematics in Year 9. The tests are compulsory. There are also national tests in these subjects that can be used by the school at the end of the fifth year. In upper secondary school there are national tests in Swedish/Swedish as a second language, English and mathematics. The tests are compulsory and results are part of the basis for student marks. There are also voluntary national course assessments in French, German, physics and biology. School leaving certificates include a record of marks for all courses in upper secondary education. Graduates of all of the 3-year upper secondary programmes are eligible to study at higher education institutions.	Schools are responsible for maintaining and improving the quality of teaching. Municipalities are required to have action plans describing how national and municipal goals are to be achieved, and to monitor the plans.	The National Agency for Education is responsible for evaluation, follow-up and school development at the national level. The **Swedish School Inspectorate is** responsible for supervision and thematic quality evaluation. Municipal and county councils appoint one or more committees to ensure that educational activities comply state regulations and guidelines as well as central education standards. (UNESCO, 2007)
SWITZERLAND	Students take matriculation examinations in five disciplines. Grades from the last year of school are also taken into consideration.		(UNESCO, 2007)
TURKEY	There are no national programme for assessing and monitoring pupils' and student achievement The Ministry of Education has established regulations regarding the number and periodicity of examinations in schools. For example, in the first three years of primary school, students are assessed on their classroom performances. In Years 4 and 5, schools may set a maximum of 2 written examinations. In Years 6 to 8, schools set a minimum of 2 written and 1 oral examination for each subject projects, assignments, on-the-job training, classroom performances, and extracurricular activities. Teachers' observations of student behaviour are also considered.	No official requirements for school self-evaluation reported.	There is an Inspection office. No details of its mandate or approach reported. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
UNITED KINGDOM	England and Wales have national tests for 7-, 11- and 14-year-olds in English and mathematics. There are assessments in science (although there are no science tests for 7-year-olds). Teacher assessments have equal status with the national assessments. In England, the test results are published for each child, each school, and the national averages (for comparison purposes). In Wales, the results are published nationally and for each local authority. In England, Wales and Northern Ireland, the General Certificate of Secondary Education (GCSE) is the principal examination taken by secondary school pupils at 16 years of age. The GCSE follows nationally agreed criteria. In England, all GCSE syllabi must be formally approved by the Qualifications and Curriculum Authority (QCA). In Wales, the syllabi must be approved by the Qualifications, Curriculum and Assessment Authority. Teachers' school-based assessments of coursework may also be a significant percentage of the final result. In Northern Ireland, assessment for Key Stages 1 to 3 are being by standardized annual reports or 'pupil profiles' In Northern Ireland, Education and Training Inspectorate (ETI), within the Department of Education, carries out inspections. The Qualification and Curriculum Authority in England, the Qualifications Curriculum Group in Wales and the Council for the Curriculum, Examinations and Assessment (CCEA)in Ireland provide advice on curriculum, examinations and assessment the CCEA in Northern Ireland conducts and moderates examinations.	In Scotland, schools are required to monitor and evaluate their performance, and must produce an annual "Standards and Quality", self-evaluation" report, as well as a development plan. In Northern Ireland, the Education and Training Inspectorate provide tools to support school self-evaluation. In England, there is an increasing emphasis on school self-evaluation. The central government and local authorities provide support. In addition, the new model of school inspections introduced in England in September 2005, places particular emphasis on school leadership and management, and school self-evaluation	England's Education and Inspections Act (2006) introduced a new single inspectorate (carrying on use of the Office of Her Majesty's Chief Inspector of Schools (OFSTED) name. Under the Act, local authorities have new powers to force failing and underperforming schools to federate or partner with another school for school improvement, In Scotland, Her Majesty's Inspectorate of Education (HMIE) carries out school inspections and promotes improvement. HMIE publishes reports on individual institutions and evaluations of the education system. (UNESCO, 2007)

	STUDENT ASSESSMENT (EXTERNAL SUMMATIVE)	SCHOOL SELF-EVALUATION	EXTERNAL SCHOOL EVALUATION
UNITED STATES	The No Child Left Behind (NCLB) Act of 2001 requires every state to measure student progress in reading and mathematics Years 3 through 8 and at least once during Years 10 through 12. Tests of science achievement were added in 2008/08. NCLB specifies intervention in the case of low school performance. There are no national school leaving examinations. Students may take standardized tests such as the Scholastic Assessment Test (SAT) or the American College Test (ACT) for admission to university, although not all institutions rely upon these examinations. Teacher assessments of student performance are taken into consideration for upper secondary school (high school) completion. Approximately 40 states have minimum competency testing school career (Years 3 or 4; Years 6, 8 or 9; and Years 11 or 12) defining "adequate yearly progress" (AYP) for all public school students and specified subgroups.	No official federal or state policies requiring school self-evaluation are noted.	Every state has developed benchmarks to measure progress toward the NCLB goal of having every child achieve proficiency (as measured by state educational standards) by the end of the 2013/14 school year. The National Center for Education Statistics (NCES) provides data to monitor student achievement and the progress of reforms. No official federal or state school inspection systems. Targeted program evaluations; audit of resource use, compliance with regulations – but not focused on teaching and learning processes. (UNESCO, 2006)

ANNEX 2: CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)

The information contained in this annex was drawn from the OECD's 2005 publication, *Formative Assessment: Improving Learning in Secondary Classrooms*, UNESCO International Bureau of Education's "World Data on Education" (WDE) reports for 2006 and Eurydice's country profile reports. The full WDE reports are available at http://www.ibe.unesco.org/Countries/WDE/2006/index.html and Eurydice's Eurybase country profile reports are available at http://eacea.ec.europa.eu/education/eurydice/eurybase_en.php.

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
AUSTRALIA	Assessment in Australia is guided in part by its approach to standards, which are based on learners' progress on a developmental continuum and performance at a particular point in time, rather than 'pass' or 'fail'. The approach is intended to assist teachers to integrate teaching, learning and assessment in classroom activities. Teachers are encouraged to use a range of assessment methods and tools to ensure a balanced judgement of student achievements over time. Outcomes-based approach is seen as a way to enhance school planning and system-wide assessment of success in key learning areas. Some states and territories provide illustrative support material to guide teaching, learning and assessment (OECD, 2005; UNESCO, 2006)
AUSTRIA	Teacher-generated assessment is based on classroom participation and the results of oral, written, practical and graphical work. Primary school pupils must sit for written exams in grade 4 (school tests) in German and Mathematics. Pupils receive end-of-term and end-of-year reports. Pupils who pass the school-leaving examination at the end of secondary higher academic school receive a matriculation certificate *(Reifeprüfungszeugnis)* (UNESCO, 2007).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
BELGIUM (FR-SPKG)	The 1997 Decree on the missions of school defines competencies and the preparation of teaching tools and assessment. It requires use of formative assessment and differentiated pedagogy. Students are to achieve equal results from education. Schools have freedom to decide the type of assessment tools they will use, and how they will communicate results. Teachers may refer to the definition of competencies as a guide for classroom-based assessment. An steering unit organizes assessments at the beginning of the 3rd and 5th years in primary school (Eurydice, 2008/09).
BELGIUM (FLEMISH COMMTY)	In Flemish-speaking Belgium, schools develop their own tests and systems to observe and monitor student progress. They may also use tests developed by their network's umbrella organization. According to the World Data on Education report, schools in Flanders also place an emphasis on "continuous assessment", on the basis of daily work in the classroom and homework. At the primary level, teachers (in consultation with the school head), make decisions as to whether or not a child will move to the next year. At the secondary level, end-of-year assessment of student achievement is decided a the school level in the context of an educational staff meeting. Teachers organize exams, under the responsibility of the "organizing powers" (*i.e.* the government, or the natural or legal person who takes responsibility for it) (UNESCO, 2007).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
CANADA	Assessment policies vary by province. Formative assessment is a prominent strategy across the different provinces, but policy supports vary.
CZECH REPUBLIC	Pupils are assessed by teachers on the basis of written and oral work and homework on a 5-point scale. The results of continuing assessment are summarized in a report at the end of each semester. Verbal assessment is authorized at all educational levels since 2005. Upper secondary schools use both continuous assessment and final assessment of pupils in a school report. The results of a pupil's education may be expressed by a mark, a verbal assessment, or by a combination of both All upper secondary schools organize their own final examination (UNESCO, 2007).
DENMARK	Denmark promotes continuous assessment to guide planning and adaptation of further instruction, to ensure that the level of teaching is appropriate, and to provide students with detailed guidance on study methods. Teachers and students determine the form and content of instruction, as well as assessment methods. The Act governing the Danish *Folkeskoler* system requires schools to conduct comprehensive and varied assessments. Assessments are to be integrated into teaching. Students are to be active participants in the assessment process (OECD, 2005).
FINLAND	Finland supports "continuous assessment", which is based on each student's progress. The aim of assessment is to encourage pupils to set their own goals and to make independent choices. Assessment should be an integrated part of daily school activities Diverse assessments, including verbal feedback, assessment interviews, and portfolio assessments, are based on objectives defined in the curriculum. The Finnish National Board of Education issues national criteria for teachers' assessment of students. Curriculum guidelines outline the principles of student assessment, *e.g.,* encouragement of student self-assessment skills. Students receive written reports at least once during the year, and at the end of the school year (OECD, 2005).
FRANCE	National diagnostic assessment protocols for French and mathematics were implemented in 2007-08 school year. The purpose is to identify students who may benefit from a more personalised teaching programme. Teachers organise regular assessments during each cycle; each student has a school record book, which is used throughout compulsory schooling. The record book is intended to ensure continuity in the transition from primary to secondary schooling. Student progression for each cycle is based on the recommendation of the teacher and the teacher council. Students receive marks as well as comments on their work and progress. At some point during lower secondary schools, each student is given an assessment to guide his/her future work in the school system (Eurydice, 2008/09).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
GERMANY	In primary education, Germany promotes alternative forms of learning and assessment. Teachers continuously assess student learning processes, performance, working and social behaviour through verbal and written controls. Continuous assessment based on written examinations and oral contributions at al levels. Assessment is teacher-led in most cases (UNESCO, 2007).
GREECE	In Greece, teachers in primary and secondary schools are responsible for assessing their students and for modifying teaching appropriately. In the Unified Upper Secondary schools, student assessments are regulated by Presidential Decree No. 86/2001 (amended in 2002). Teachers are encouraged to use a variety of assessment approaches and techniques with the aim of fostering students' self-knowledge, and keeping parents/guardians fully informed. Forms of assessment include: diagnostic assessment, oral feedback, composite creative projects, assessment of assignments and activities that make up the student's optional performance and activity file, and marks on written examinations for promotion or graduation (UNESCO, 2007).
HUNGARY	In Hungary, teachers make regular assessments of pupil performance and progress during the school year. The end-of-term and end-of-year marks are based on continuous assessment. Assessments may be written, verbal, or sometimes based on tests. From Eurydice: The performance and progress of pupils are regularly evaluated by teachers throughout the school year on the basis of principles set in the local curriculum. Pupils are generally assessed based upon the traditional numeric grading (scale 1-5). It is to be expressed in a written statement whether the pupil has done excellently, well or satisfactorily or needs coaching in the middle or at the end of the academic year in grades 1-3 and at the middle of the academic year in grade 4. In the first three grades pupils cannot be forced to repeat a year. If a pupil is assessed as one who needs coaching, the school also has to involve the parents of the pupil in the evaluation and reveal the factors impeding progress and has to put up a proposal as regards the necessary measures to cease them. In grade four and above pupils may be made to repeat a year. The pedagogical programme of schools may stipulate to use descriptive assessment or other grading instead of using marks (scale 1-5) at mid-term, at the end of the school year and during the year. (UNESCO, 2007)

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
ICELAND	Iceland describes the purpose of school and teacher assessment as being to, first, check the effectiveness of teaching and learning, and to provide both students and their parents with information on their progress. Assessment is not standardized between schools and teachers. Student progress reports may be given in the form of marks (numbers or letters), or verbal or written descriptions, and are made at regular intervals during the year, and at the end of the school-year. Examinations and other forms of assessment, usually written, are carried out by individual teachers and schools. Assessment is therefore not standardised between different schools and teachers. The way in which the reports on pupils' progress are written varies greatly: the assessment can be in the form of a number, a letter or a description either oral or written. Reports are given at regular intervals throughout the school year and at the end of each year. The purpose of assessment by the school and the teacher is above all to help improve learning and teaching and to provide both the parents and the children with information on how their studies are progressing (UNESCO, 2007).
IRELAND	Ireland's recently revised curriculum places student assessment at the centre of the teaching and learning process. Assessment strategies are to be used to identify and provide for student needs. Ireland also places an emphasis on early diagnosis of serious literacy and numeracy problems at the beginning of primary schooling. Teachers and learners assess progress, and use information to shape next steps. In primary school, assessment may include observation, teacher-designed task and tests, conferencing and portfolio assessment. In secondary schools, oral assessments in languages and hands-on assessments in subjects like geography and science are increasingly common. Schools have considerable autonomy in deciding the teaching and assessment methods. The revised curriculum emphasizes student learning styles, and the importance of integrating assessment into all areas of teaching and learning. Primary teachers have received extensive in-service training with adoption of the new curriculum. The curriculum also includes guidance on appropriate assessment procedures (UNESCO, 2007).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
ITALY	Italy introduced new criteria for pupil assessment in 1994/95, concluding an experimental phase that had begun in 1977 with the adoption of Law No. 517. The law replaced the traditional grading system with teachers' analytic and synthetic assessments. Schools prepare a plan (the *Piano dell'Offerta Formativa*, POF), reflecting the local context and needs. The POF covers details regarding the teaching staff, timetable, and on individualized study plans. Italy's 2004 school reform introduced the *individual skills portfolio.* The portfolio records the student's progress toward educational targets specified in the individual study plan. It includes teacher and parent comments (and where appropriate, student comments). It also includes teacher and family remarks on the teaching methods, on students' personal work and projects, summaries of discussions with students and/or parents, the results of tests, and comments based on systematic observation. Students are continuously assessed. At the end of lower secondary school, students take a national school-leaving examination. Upper secondary school students are assessed on the basis of examination results, as well as participation in school activities in general, initial preparation and subsequent progress, and other information obtained from contacts with the family (OECD, 2005).
JAPAN	Japan has a strong emphasis on promoting Lifelong Learning, but country reports do not indicate any strong emphasis on policies to support formative assessment.
KOREA	Korea's 1995 Presidential Commission on Education Reform established individual comprehensive personal records, with the aim of supporting diagnostic, formative and summative assessments of each student's academic achievement and social development. The information is to be used to improve the teaching and learning process for each student. High schools (vocational, science and special purpose) also use the information to select students; colleges and universities use the personal records along with the College Scholastic Ability Test to select students. The Korea Institute of Curriculum and Evaluation (KICE) has developed a standards for criterion-referenced assessment in each curricular subject (UNESCO, 2006).
LUXEM-BOURG	Teachers in primary schools conduct ongoing assessment, as well as overall summary assessment (generally written tests). In 1996, entrance examinations were replaced by standardised and psychological testing, and students and their parents receive recommendations regarding the academic path to follow. Teachers at secondary schools organize up to three tests a term in each subject (Eurydice, 2008).
MEXICO	Information not available in English or French.

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
NETHER-LANDS	In the Netherlands, teachers keep records of individual students' progress, with the results of oral and written tests and projects. Schools may use tests (often standardized) to assess student progress. There are also national tests. In many cases, schools are able to compare their results with other schools – to highlight weak areas, and adjust teaching. Most school issue progress reports three times a year. Schools choose whether they will give marks or indicate student achievement in another way (*e.g.* descriptive). Many schools have adopted monitoring systems to record individual student progress in a systematic way, and to better identify individual student needs and adapt teaching.
NEW ZEALAND	At the national level, the New Zealand Curriculum provides guidelines on teaching, learning and assessment for all students in all schools. Teachers develop assessment plans based on objectives spelled out in the national curriculum, and develop their own school curriculum, classroom programmes and assessment plans based on the national curriculum. Teachers may adapt programmes according to the needs of their students. There are teacher development programmes for new curriculum statements. The government provides tools to support assessment through the asTTle (assessment tools for teaching and learning, now available in a 4th version). The government has also published exemplars focusing on curriculum and formative assessment principles. The government has published exemplars focusing on curriculum and formative assessment principles. Most recently, the National Education Monitoring Programme (NEMP) has been modified to include information useful to implementation of the new *National Standards*, which are based on "assessment for learning" principle. In New Zealand, primary level student assessments are based on teachers' qualitative judgments, of student performance and progress. There are no national tests. At the secondary level, the National Certificate of Educational Achievement (NCEA) sets standards for student performance. Students are evaluated against written criteria, which are accompanied by exemplars showing expected levels of student performance. Since 2008, Maori assessment experts have been developing assessment tools to be used in Maori-medium settings (see NZ Ministry of Education (2010), http://www.minedu.govt.nz/~/media/MinEdu/Files/TheMinistry/AssessmentPositionPaperSep2010.pdf
NORWAY	In Norway, assessment is seen as a tool for promoting the student's learning and development. Students should play an active role in the process, and also develop skills for self assessment. Assessments (unmarked) are an integral part of the daily learning process. The results of daily assessments are included in regular conferences between teachers, students and parents. Students do not receive marks at all during primary school. Marks are introduced in lower secondary schools as part of student assessment. During primary school (Years 1 to 7) there are no formal assessments of pupils. At lower secondary stage (Years 8 to10), teachers award marks for each subject twice a year. In upper secondary school, teachers conduct continuous assessment, and students sit end-of-year examinations (UNESCO, 2007).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
POLAND	The 1995 amendments to educational legislation introduced core curriculum, and criteria for school-based assessment and examinations. The central authority on evaluation advises that assessment should be used as an instrument for managing learning, and not as a tool for selection. There is an increasing emphasis on assessment of skills rather than knowledge and facts (UNESCO, 2007).
PORTUGAL	At the beginning of the school year, the pedagogical council, in line with national curriculum guidelines, defines the assessment criteria for each cycle and year of schooling, as proposed by the teachers' council, in the 1st cycle, and by curricular departments and cycle co-ordinators in the 2nd and 3rd cycles. Assessments are diagnostic assessment, formative and summative (Eurydice, 2006/07).
SLOVAK REPUBLIC	Students are continuously assessed on the basis of written and oral work. Students in primary school receive official assessments twice a year, using a 5-point marking scale (1 = excellent; 5 = failed). Assessments are organised throughout the school year (written and oral tests and homework). Students in first grade receive verbal assessments. Since 1995/96, students in grades 2 to 4 may also receive verbal assessments, if parents and the pedagogical board of the school agree. At the end of the ninth grade of primary school *(základná skola)*, students are awarded on the basis of their school results (there is no final examination). Assessment in upper secondary school follows the same guidelines as for compulsory levels (UNESCO, 2007).
SPAIN	In Primary Education, teachers assess pupils' progress in all areas with a global and continuous approach. Teachers are responsible for the decisions on their promotion, taking special account of the information and criteria of the class teacher. Promotion is automatic within the same cycle of Primary Education but progression from one cycle to the next is contingent upon meeting the curricular aims for that particular cycle. A pupil may repeat a year, but only once in the primary level. Pupils who continue to the next cycle, but who are negatively assessed in one or more areas, must receive appropriate support to help them catch up. Likewise, special attention is paid to the early detection of learning difficulties and to the prevention of school failure at an early age. An official academic certificate is not awarded at the end of primary education, but it is awarded at the end of the basic education, which includes Primary and Lower Secondary Education. In Lower Secondary Education, assessment is continuous (*i.e.* integrated into the teaching and learning process) and separate for each subject. Remedial measures are adopted within the process of continuous assessment when students are not progressing at an appropriate rate. All decisions regarding student assessment are made jointly by the teaching team (within the framework established by education authorities). At the end of each year, all the teachers of the group jointly decide on each pupil's promotion after considering the attainment of the objectives of the year. Pupils may take a special examination in subjects they have not passed at the end of the school year. Assessment is continuous in upper secondary school and differentiated according to subjects. Students are assessed by the teaching team, (coordinated by the form teacher) under the advice of the Counselling Board of the educational institution. Teachers also assess the teaching-learning processes and their own teaching practices (Eurydice, 2009/10).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
SWEDEN	Sweden reformed curricula and standards in the 1990s, and introduced a "goal and results-oriented" system for the administration of education. The curriculum stresses the importance of "learning-to-learn" and of students taking responsibility for their own learning. Teachers practice continuous assessment throughout compulsory education. Teachers use local action plans and grading criteria to assess student learning. In the mid-1990s, teacher trade unions and trade associations agreed on a preference for common planning, to compensate for the fact that the teaching load was no longer regulated at the national level. Teachers organize work teams to better deal with the increase in student options and individualized reporting forms. The national government has provided financial support for professional development and exchange of experience in networks of school administrators, teachers and other personnel. The government has also published written guidance on the intentions of the reform via the Internet, and holds workshops, conferences and seminars for various target groups. Grades are introduced in the eighth school year and are awarded on a three-point scale: Pass, Pass with Distinction and Pass with Special Distinction. Students who do not achieve the goals of a certain subject receive a written assessment instead of a grade. Pupils automatically move to a higher class each year. A school leaving certificate is awarded to students who successfully complete the final year In upper secondary school, assessment is continuous with marks awarded at completion of each course. There are national tests in certain subjects. There is no final examination, but students receive a leaving certificate. The certificate includes a summary of the student's coursework and grades received. Students may re-sit to improve grades if they wish (UNESCO, 2007).
TURKEY	Turkey introduced 'process-based assessment' in 2006. Teachers assess primary school student performance during the school year based on projects, exam scores, homework, classroom participation, attendance, behaviour, and so on. In upper secondary school, progression to the next grade is based on students' achievement across courses, or the average annual level of attainment. Individual teachers assess performance through written or oral examinations, practical examinations, homework, and projects. The average score for a course in a semester is calculated on the basis of the average of all course marks obtained during that semester. Successful students progress to the next grade (UNESCO, 2007).
UNITED KINGDOM	In England, "key thinking skills" for information processing, reasoning, enquiry, creative thinking and evaluation are embedded in the national curriculum. In England and Wales, the results of national assessments are combined with teachers' assessment at the end of each key stage. The Assessment for Learning (AFL) programme supports formative assessment practice. In Scotland, the curriculum is integrated with assessment; it is recognised that students learn at different rates. Scotland also supports the Assessment is for Learning (AiFL) programme (UNESCO, 2007).

COUNTRY	CLASSROOM-BASED ASSESSMENT (FORMATIVE AND SUMMATIVE)
UNITED STATES	No federal policies related to classroom-based formative or summative assessment (UNESCO, 2006).

ANNEX 3: OECD COUNTRY POLICIES ON ASSESSMENT OF TEACHER PERFORMANCE

The information contained in this annex was drawn from the UNESCO International Bureau of Education's "World Data on Education" (WDE) reports for 2006 and Eurydice's country profile reports. The full WDE reports are available at http://www.ibe.unesco.org/Countries/WDE/2006/index.html and Eurydice's Eurybase country profile reports are available at http://eacea.ec.europa.eu/education/eurydice/eurybase_en.php.

COUNTRY	*APPRAISAL OF TEACHER PERFORMANCE*
AUSTRIA	No information on teacher appraisal.
AUSTRALIA	No information on teacher appraisal.
BELGIUM (FLEMISH COMMUNITY)	Flanders sets out principles for teacher assessment. The assessment process is to be used as a positive process, and should be based on ongoing work. Criteria for assessment are based on individualized job descriptions, which are mandatory for all staff. Assessments must be conducted at least every four years. Each staff member has two evaluators – with the first evaluator being responsible for guidance and coaching. The government recommends and provides funding for training of evaluators.
BELGIUM (FRENCH-SPEAKING COMMUNITY)	In the French-speaking Community of Belgium, teachers belong to networks. Management personnel in the networks are responsible for developing teacher assessments. Inspectors also play a role in assessment of teachers, upon the request of a school head.
CANADA	No information on teacher appraisal.
CZECH REPUBLIC	School heads, who are responsible for quality of education, assess teacher performance. There are no centrally-set criteria or methods. Additional methods such as self- and peer-assessment and assessment by students and parents are also being encouraged.
DENMARK	There is no formal evaluation of teachers once they have passed a two-year probation period (with the exception of Folkeskole teachers).
FINLAND	Teachers are not formally evaluated. However, most schools have quality systems, which include annual development discussion and appraisals.
FRANCE	National inspectors have primary responsibility for assessment of teacher performance. The inspector gives every teacher a mark, based on educational and administrative criteria. Teachers are assessed approximately once every four years. They may also request an assessment to advance their careers. At the primary level, there is approximately 1 inspector for 350 teachers; at the secondary level, the ratio is 400 to 1.
GERMANY	Teachers are assessed before changes in their civil servant status (at regular intervals). Ministers of Education and Cultural Affairs in the Länder set out appraisal guidelines for assessing teachers. Teacher appraisals must cite the assessment criteria and use of assessment.
GREECE	Evaluation of the educational system is linked mainly to evaluation of teachers and students. Data collection focuses on input, processes – including pedagogical practices – and results.
HUNGARY	There is no requirement for assessment of teachers, but many individual institutions have developed their own performance assessments in line with the Public Education Act, and usually include formative assessment of teachers.
ICELAND	Individual teachers are not appraised.
IRELAND	Two modes of evaluation are reported: Inspectors may evaluate and report on teachers' work (most commonly practiced in primary level. At the post-primary level, inspections focus on whole school evaluations.
ITALY	Teachers are assessed at the end of the initial induction period, if a permanent teacher requires an assessment, and in the context of a disciplinary procedure or release of service due to poor performance.

JAPAN	No information on teacher appraisal.
KOREA	No information on teacher appraisal.
LUXEMBOURG	No information on teacher appraisal.
MEXICO	No information on teacher appraisal.
NETHERLANDS	School boards are responsible for recruiting, training and assessing staff. Teachers are assessed in job performance interviews with school heads (usually bi-annual). The assessment includes classroom observations. In secondary schools, teachers' peers and students may also be consulted. Some schools conduct annual assessment interviews, which are separate from the job interview. The assessments cover performance, as well attitudes toward colleagues and professional development.
NEW ZEALAND	No information on teacher appraisal.
NORWAY	Individual teachers are not appraised.
POLAND	Teacher performance are initiated by school heads but may also be conducted on the request of the kurator, school council or parent council. The assessment may incorporate views of student government. Teachers receiving a negative assessment may participate in further training, and request a follow-up assessment.
PORTUGAL	Teachers submit a "critical reflection" certificates of professional development since the last assessment. Assessments occur after a certain number of years based on where they are in the career scale.
SLOVAKIA	Since 2003, school leaders have been required to conduct annual assessments of educational and special employees. Assessments are based on methods developed for school inspections, focused on educational processes. Prior to 2003, teachers were assessed only at the end of the induction period.
SPAIN	Education authorities of the Autonomous Communities are responsible for assessment of teachers. The plans must be publicly announced, and must outline the criteria and methods of assessment.
SWEDEN	There are no legal requirements for teacher assessment. However, school heads hold regular individual development dialogues with teachers.
TURKEY	School leaders approve teachers' annual plans, monitoring, and identifying and addressing weaknesses. They are also responsible for teachers' professional development.
UK	In *England*, revised guidelines on annual teacher performance assessments were issued in 2007. Schools must develop pay and performance management policies, which, among other requirements, must link teacher performance with plans for school improvement and school self-valuation, include classroom observations, provide for training as needs arise. In *Wales*, revised guidelines were issued in 2002. The governing body of a school is required to establish the performance management policy of the school, and to assess teachers annually. The policies must establish the performance objectives and monitoring process. In *Northern Ireland*, teachers are reviewed annually, usually by an individual with management or curricular responsibility for the teacher. The assessments are based on two classroom observations as well as review of objectives set out at the beginning of a period, and cover areas of practice, professional development, student and curriculum development. The reviews should also link to the school development plan. Teachers in *Scotland* are not assessed individually.
US	The No Child Left Behind Act defines the qualifications needed by teachers and paraprofessionals who work on any facet of classroom instruction. It requires that states develop plans to achieve the goal that all teachers of core academic subjects be highly qualified by the end of the 2005/06 school year. No information on teacher appraisal (policies vary by state).

THE OECD EDUCATION WORKING PAPERS SERIES ON LINE

The OECD Education Working Papers Series may be found at:

- The OECD Directorate for Education website: www.oecd.org/edu/workingpapers
- The OECD's online library, www.oecd-ilibrary.org/papers
- The Research Papers in Economics (RePEc) website: www.repec.org

If you wish to be informed about the release of new OECD Education working papers, please:

- Go to www.oecd.org
- Click on "My OECD"
- Sign up and create an account with "My OECD"
- Select "Education" as one of your favourite themes
- Choose "OECD Education Working Papers" as one of the newsletters you would like to receive

For further information on the OECD Education Working Papers Series, please write to: edu.contact@oecd.org.