

OECD Education Working Papers No. 42

School Evaluation: Current Practices in OECD Countries and a Literature Review

Violaine Faubert

https://dx.doi.org/10.1787/218816547156

OECD

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

15-Dec-2009

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2009)21
Unclassified

School Evaluation: Current Practices in OECD Countries and a Literature Review

OECD Education Working Paper No. 42

By Violaine Faubert

This paper was prepared for the OECD by Violaine Faubert, a graduate student at the Institut d'Etudes Politiques de Paris (Sciences Po) in the context of the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.

Contact: Mr. Paulo Santiago [Tel: +33(0) 1 45 24 84 19; e-mail: paulo.santiago@oecd.org]

JT03276248

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.

---

www.oecd.org/edu/workingpapers

Applications for permission to reproduce or translate
all or part of this material should be made to:

Head of Publications Service
OECD
2, rue André-Pascal
75775 Paris, CEDEX 16
France

Copyright OECD 2009

ABSTRACT

This paper examines the current academic and policy literatures concerning school evaluation in primary and secondary education within the OECD countries. First, it provides a typology of the existing systems of school evaluation across the OECD. It encompasses the diverse criteria and instruments commonly used to carry out schools evaluation, as well as the players involved in the design and implementation of school evaluation. It also describes potential consequences for schools. Second, this paper analyses how school evaluation schemes are interrelated with other components of the evaluation framework, such as teacher evaluation and system evaluation. The potential complementarities, duplication and inconsistency of objectives stemming from these interrelations are discussed. Third, this paper presents the advantages and drawbacks of different approaches to school evaluation, the resistance and implementation difficulties resulting from misalignment of interests between different stakeholders, and possible ways to overcome impediments to implementation. Finally, it reviews the quantitative and qualitative evidence available on the impact of different school evaluation schemes on school performance, student learning and the incentives for the teaching staff. It concludes by considering the circumstances under which school evaluation schemes seem to be more conducive to school improvement. The effectiveness of school evaluation schemes relies on developing competencies for evaluation and for using feedbacks. Alignment of stakeholders' interests is also critical to have the support of those being assessed.

RÉSUMÉ

Cet article dresse une revue de la littérature académique et empirique consacrée à l'évaluation des établissements scolaires primaires et secondaires au sein des pays de l'OCDE. Une typologie des différents systèmes d'évaluation existants est tout d'abord esquissée. Les différents instruments et critères utilisés pour procéder à l'évaluation des établissements scolaires, ainsi que les acteurs impliqués dans la conception et la mise en œuvre des systèmes d'évaluation des écoles, sont précisés. Les conséquences induites par les résultats de l'évaluation pour les établissements sont également évoquées. Cette revue de littérature analyse ensuite les relations existant entre l'évaluation des établissement scolaires et autres composantes du cadre d'évaluation du système scolaire, notamment l'évaluation des enseignants et l'évaluation du système éducatif. Les sources de complémentarité, effets de doublon et possibles incompatibilités entre ces différentes composantes, sont évoqués. Les avantages et les inconvénients des différentes approches de l'évaluation des écoles, ainsi que les résistances et les difficultés rencontrées dans leur mise en œuvre en raison des divergences d'intérêt et de point de vue des différents acteurs (autorités éducatives, inspecteurs, chefs d'établissement, corps enseignant…), sont soulignés. Les résultats quantitatifs et qualitatifs d'études empiriques destinées à tester l'impact de l'évaluation sur les performances des élèves, la gestion des établissements, la motivation des enseignants et, plus largement, sur les incitations des agents, sont mentionnés. Cette revue de littérature décrit enfin les environnements les plus favorables au développement des mécanismes d'évaluation les plus efficaces. Ceux-ci supposent notamment le développement de compétences en vue de procéder à des évaluations et d'en exploiter les résultats, ainsi que le soutien et la convergence de vues des acteurs impliqués dans le processus.

TABLE OF CONTENTS

1. INTRODUCTION 6

2. KEY FEATURES OF SCHOOL EVALUATION SCHEMES IN OECD COUNTRIES: A FRAMEWORK FOR A COMPARATIVE INVESTIGATION 7

2.1 Purposes and scope of school evaluation 7
2.1.1 Purposes of school evaluation 7
2.1.2 Scope of school evaluation 8
2.1.3 Users of results from school evaluation 10
2.2 Responsibilities for school evaluation 11
2.2.1 Agencies with responsibility for school evaluation 11
2.2.2 Other actors involved in the conception and implementation of school evaluation 13
2.2.3 External evaluators: qualifications and training 14
2.3 Evaluation procedures 16
2.3.1 Reference standards and aspects assessed 16
2.3.2 Evaluation instruments 19
2.3.3 Methodology and procedures 20
2.4 Using evaluation results 22
2.4.1 Using evaluation results for accountability 22
2.4.2 Using evaluation results for improvement 24
2.4.3 Developing competencies to use evaluation results effectively 26

3. ARTICULATION BETWEEN SCHOOL EVALUATION AND OTHER TYPES OF EVALUATION 28

3.1 Use of school evaluation results for the evaluation of the school system 28
3.1.1. Use made of external school evaluation results 28
3.1.2. Use made of findings from internal evaluations 29
3.2 Articulation between school evaluation and teacher evaluation 29
3.2.1 The individual evaluation of teachers and the collective responsibility for school teaching quality 29
3.2.2 The evaluation of teaching as part of schools' evaluation 30
3.2.3 Use made of teacher evaluation results for school evaluation 31

4. ISSUES IN DESIGNING AND IMPLEMENTING SCHOOL EVALUATION APPROACHES 32

4.1 Advantages and disadvantages of approaches to school evaluation 32
4.1.1External school evaluation 32
4.1.2 Use of student results to evaluate schools 33
4.1.3 Performance-based funding 34
4.1.4 Public release of results and school “rankings” 34
4.2 Difficulties in implementing school evaluation 36
4.2.1 Conflicting perceptions of school evaluation 36
4.2.2 Implementation difficulties 37
4.2.3 Lack of appropriate information and training to guide evaluation and improvement 39

4.2.4 Ways to overcome obstacles 40
5. EMPIRICAL EVIDENCE ON THE EFFECTS OF SCHOOL EVALUATION 43
5.1 Difficulties in measuring the impact of school evaluation 43
5.2 Empirical evidence on school evaluation schemes 43
5.2.1 Empirical evidence on the impact of external evaluation systems on school performance 43
5.2.2 Empirical evidence on the impact of evaluation schemes for student performance 45
5.2.3 Effects of school evaluation on teachers and school leaders 46
6. CONCLUDING REMARKS 48
REFERENCES 49

1. INTRODUCTION

1. Public demands for educational accountability and improvement intensified in most OECD countries over the past decades. Increased demands for effectiveness, equity and quality in education are justified by the necessity to cope with new economic and social needs. On the one hand, improving the quality of education is necessary in order to meet the ever increasing demand for skilled workers in a context of global economic competition. On the other hand, accountability systems aim at ensuring that school objectives are being met with an effective use of resources.

2. Although evaluation within school systems is not a recent concern, it used to be limited to the evaluation of students for many years. The systematic evaluation of teachers, programs, or the school as a whole, has emerged more recently, and is still far from being common practice in even the most advanced education systems (Nevo, 1998). Evaluation schemes are now increasingly being considered as potential levers of change that could assist with decision-making, resource allocation or school improvement.

3. There is a wide diversity among OECD countries in the ways they approach the evaluation of schools. Scope and methods of school evaluation, criteria and standards used and data gathering instruments differ largely across education systems, according to the educational context and the stakeholders involved in the design and implementation of school evaluation schemes. This variety is the result of different traditions, diverse administration systems and educational policies, not only related to evaluation, but also to school autonomy. Considering the varied degrees to which decentralization policies are implemented in the different countries, the roles and functions of school evaluation can vary (Scheerens *et al*., 1999).

4. Though school evaluation systems vary in their characteristics, they share a common global purpose of improving teaching and learning. This report reviews the current state-of-the-art: national approaches are compared in order to assess the pros and cons of the various evaluation schemes in place. However, assessing the comparative effectiveness of approaches to school evaluation is challenging, since school systems differ across countries in a wide range of dimensions.

5. This report has five further sections. Section two aims at reviewing approaches to school evaluation in OECD countries. The links between school evaluation and other components of the school system evaluation framework are investigated in section three. Section four analyses the advantages and drawbacks of different approaches to school evaluation. Section five presents empirical evidence on the impact of school evaluation schemes upon education outcomes. Finally, section six offers some concluding remarks.

2. KEY FEATURES OF SCHOOL EVALUATION SCHEMES IN OECD COUNTRIES: A FRAMEWORK FOR A COMPARATIVE INVESTIGATION

6. This section describes the key dimensions of school evaluation within school systems of OECD countries.

2.1 Purposes and scope of school evaluation

2.1.1 Purposes of school evaluation

School evaluation serves two main purposes: improvement and accountability

7. *School evaluation serves two interlinked purposes*, improvement and accountability. *School improvement* relates to access to education (equity) and education performance (quality and efficiency). School evaluation for improvement aims at closing achievement gaps between low-performing and high performing schools, as well as to enhance the performance of all students. School evaluation for improvement purposes generally implies a formative approach. However, definitions and perspectives regarding the purpose and the focus of improvement can vary according to different stakeholders.

8. *School accountability* aims at providing information to policy makers and the public about value for money, compliance with standards and regulation and quality of the services provided. School evaluation for accountability purposes generally implies the use of summative approaches. Conceptually speaking, three dimensions of accountability are particularly relevant for school evaluation. First, contractual accountability is externally directed, and focuses on meeting the requirements of the school system and contributing to improve its quality. Second, moral accountability is focussed on meeting the needs of parents and students. Third, professional accountability is focussed on meeting one's expectation and those of colleagues and is more internally directed (Learmonth, 2000; Gurr, 2007). Potential tensions may exist between these forms of accountability. In practice, various levels of accountability may co-exist. First, a traditional type of vertical (hierarchical) external accountability, in which schools provide information to the relevant public authorities (local or national), is present in most countries. Similarly, schools may provide their supervisory boards with insight into the adequacy of their management and practices. In this context, external evaluation plays a major role. Second, a horizontal type of accountability, in which schools provide their community and stakeholders with insight into their processes, choices and results has emerged more recently, alongside with the development of internal evaluation. In practice, schools are increasingly held accountable to multiple levels of educational authorities and to the wider community in which they are involved (parents, pupils).

School evaluation gains prominence as schools are granted more autonomy and market forms of accountability are developed

9. There is an increased prominence of school evaluation as school systems decentralize with further autonomy given to individual schools and as market forms of accountability gain in importance. There has been a general international trend to implement devolution of responsibilities for resource management, curriculum development, and school evaluation to the school level. This increased autonomy has been balanced by the strengthening of accountability through the setting of outcome assessments and national standards that all schools should meet. In this context, strong leadership is needed to report to central authorities and implement goal-oriented changes at the local level. Within a context of large levels of autonomy, the evaluation function itself could be given to the school with external evaluation taking the shape of inspecting school approaches to evaluation. An example would be the situation in the Netherlands

where schools are required to carry out school self-evaluations, while the inspectorate or external review panels build further on these self evaluations (Scheerens *et al*., 1999).

10. The second major trend is the shift from a bureaucratic to a market-driven accountability system. The shift in accountability takes place when parents are offered wider choice in the education of their children. Several countries have attempted to raise school quality by enhancing parental choice and allowing schools to compete for students. In some countries, such as the Netherlands and the United Kingdom, accountability is sought in the form of strengthening consumer control (league tables, school guides) more than in the form of active measures (rewards and sanctions) from higher administrative levels. In a school system significantly relying on parental choice, external evaluation serves two purposes: to assure that schools are meeting the centrally-defined standards that justify their receipt of public funds; and to assure that parents have reliable information to assist with their decisions.

2.1.2 Scope of school evaluation

Domains and areas for evaluation

11. *Four evaluation domains can be identified* (Mc Beath *et al*., 1998): first, the evaluation of outcomes, second, the evaluation of processes at the classroom level, third, the evaluation of processes at the school level and fourth, the school environment. More precisely, outcomes are associated with academic achievement, personal and social development, and pupils' destination after leaving school. Processes at the classroom level relate to the quality of learning and teaching. Processes at school level relate to the school as a learning, social and professional place. Finally, evaluation of the school environment is focused on the relations between the school and the parents, as well as with the links between the school and the local community.

12. *Evaluators may attach importance to processes, outcomes or both.* The definition of what may be considered 'input', 'process' and 'outcome' varies across countries. Processes are a set of practices that transform an 'input' into an educational output. Processes usually relate to the management of financial, human and material resources needed to contribute to the quality of a school's performance in compliance with the regulations. The outcomes are the results achieved by a school with respect to its aims. They relate to the social and cognitive skills acquired at school and may focus on how pupils perform in their careers beyond the completion of schooling. As regards processes, evaluation covers the compliance of educational practices with the legislation, as well as the quality of practices developed by the schools for achieving the aims set by the education authorities. From the standpoint of outcomes, the performance of pupils is generally examined as regards the acquisition of cognitive and social skills. The majority of approaches to evaluation are concerned with both processes and outcomes and generally cover all school activities. By contrast, evaluations exclusively focussed on processes are usually concerned with specific aspects, such as compliance with regulations or the evaluation of internal evaluation procedures (Eurydice, 2004).

Focus of school evaluation

School evaluation focused on procedural aspects

13. Evaluation exclusively focused on processes is primarily intended to monitor compliance with regulations. It is usually concerned with aspects such as the composition of classes, the tasks and workload of teachers, the use of school infrastructure, and the management of human and budgetary resources, rather than with educational tasks. In Greece, for instance, evaluation is centred on human and functional resources administration. Great emphasis is laid on staff management: punctuality of staff, their involvement in school activities, relations between teachers and parents and collaboration between teachers

are examined. In Austria, school inspectors look into the quality of teaching and the implementation of administrative tasks. They are further in charge of examining legal compliance in all school-provision matters (e.g. construction of schools) by the communities and provinces.

14. *In a minority of countries, external evaluation consists of an evaluation of the internal evaluation processes*. This is increasingly the case in Austria. Similarly, in Iceland, external evaluators are contracted by the Ministry of Education to undertake the evaluation of the internal evaluation procedures developed by compulsory and upper secondary schools, in order to verify whether the latter comply with the goals set by the Ministry of Education.

School evaluation focused on outcomes

15. The majority of approaches to evaluation concerned with outcomes rely on the results obtained by pupils in national tests and examinations, as a means to evaluate the performance of individual schools. National and international assessments are also used to measure the extent to which students achieve national standards. In most countries, school evaluation focuses either on the levels of student performance or on gains in student performance. School performance measures generally rely on student test scores. They might focus on overall gains in student performance or rather use an adjusted version of student gains.

16. *A minority of countries have adopted an approach to evaluation chiefly concerned with pupil outcomes*. This applies to Hungary and the United States where evaluation is focussed on pupil attainment in national tests. In the United States, the *No Child Left Behind* (NCLB) Act adopted in 2001 comprises strong accountability measures designed to encourage schools to meet specific standards (US Department of Education, 2002). Test results for each school must be reported and compared to state-determined targets for adequate yearly progress (AYP) that are designed to meet the law's requirement that all students reach 'proficiency' in each subject by 2014. Hungary also adopts an approach to evaluation concerned solely with the results obtained by pupils in national tests. As standardized evaluation structures are lacking in Hungary, the first regular performance measurement of pupils that served the purposes of evaluation was the reading comprehension, and mathematical competence survey covering each pupil in grades 5 and 9. The measurements carried out with similar contents over several years since 2001 provide opportunity both at school and at system-level to assess the effectiveness of education.

17. *Assessed outcomes may relate not only to the levels of performance attained by schools, but also to equity of results within the school.* Schools are increasingly required not just to improve the educational attainment of pupils, but also to assume broader social responsibilities, including the integration of children with special educational needs, the social mix, equality of opportunity for disadvantaged pupils, and the integration of immigrant children (OECD, 2007a). Consequently, school evaluation may focus on whether schools meet equity targets. This applies to the United States and is one of the main objectives underlying the *No Child Left Behind* Act adopted in 2001 (US Department of education), which mandates that states narrow the achievement gap between their highest and lowest performing students. The NCLB foresees that for each measure of school performance, states must specify annual targets to be met by specific groups of students. The subgroup provision applies to students who are economically disadvantaged, are members of racial or ethnic minorities, are disabled, or are Limited English Proficient (LEP). Schools and districts must test students and meet targets for each student subgroup in the school in order to make "adequate yearly progress" (AYP) and obtain federal funding. The contribution of this requirement to equity is however controversial, since "schools with a sufficient number of students in each of several targeted groups to be reported are less likely to meet AYP targets than schools of the same size and similar performance but with a homogeneous student body" (Linn R., 2005).

Evaluation focussed on both outcomes and processes

18. *External evaluation is typically concerned with both processes and outcomes,* whether evaluators are provided with a pre-determined list of aspects to be assessed (as in the Flemish Community of Belgium, the Netherlands, Portugal or the United Kingdom) or not (as in France, Poland and Iceland) by educational authorities. In France, evaluation by two inspection authorities, the national education inspection (IEN, for primary schools, vocational education and apprenticeships) and the regional pedagogical inspectors (IPR for secondary general and technological education), focuses on compliance of the school plan with national legislation, its implementation and its consequences for the results of pupils. Evaluation of the school head by the *recteur* of the *académie* focuses on the processes adopted by the school in areas such as the pedagogical methods, the resources involved and the results of pupils. As regards processes, the regional inspectorate (*kuratorium*) in Poland examines compliance with national legislation in areas such as the curriculum, teachers' qualifications and their rules for assessing pupils. As regards outcomes, the *kuratorium* assesses the performance of pupils and checks whether regional educational objectives and the aims of each school are achieved. In the United Kingdom and Iceland, the assessment of processes concentrates on whether the curriculum, the composition of teaching staff, and building management comply with national legislation. Output measures considered include the results of pupils in tests or examinations (Eurydice, 2004).

Evaluation whose focus varies depending on the authorities involved

19. In certain countries in which local authorities are responsible for school evaluation, the scope of external evaluation varies depending on the entity concerned. In countries in which an approach to evaluation is decentralised to the local authorities, the content of external evaluation is determined by them. This applies to decentralized states such as Canada, Spain, and Germany. In Germany and Spain, external evaluation is concerned with both processes and results in certain Länder or Autonomous Communities, whereas in others it only covers processes. In Germany for instance, each individual Land sets the content of the external evaluation of schools. In the Land of Brandenburg, the *Schulaufsicht* solely evaluates teaching processes, while evaluation criteria can include the attainment of pupils in Bremen (Eurydice, 2004). In Spain, the Autonomous Communities are responsible for defining school evaluation plans. In the Autonomous Communities that have not yet adopted plans for evaluation, the focus of evaluation varies. As a general rule, it tends to monitor the observance of regulations, but three Autonomous Communities (Navarra, the Basque Country and Rioja) also take into account pupils' attainment. In Finland and Hungary, as there are no national regulations for the evaluation of individual schools, education providers are responsible for evaluating education (respectively, the municipalities and the local and county government) and decide about the approach to local evaluation. In Sweden, the municipalities themselves are responsible for the organisation and implementation of evaluation locally. Though the municipalities are free to determine the content of the quality appraisals that they conduct of their schools, the Swedish National Agency for Education (NAE, or Skolverket) is recommending that these appraisals deal not just with processes, but also school performance and pupil achievement.

2.1.3 Users of results from school evaluation

20. *Results from school evaluations are used by various stakeholders*. For educational authorities and parents, evaluation systems are intended as an accountability tool. As regards policy-makers, external evaluation and students' assessments provide governments with information to compare school performance within the country and in relation to other countries. Information on the implementation and achievement of the stated educational goals may be used by educational authorities to improve the allocation of resources. In market-oriented systems, external evaluation is justified in terms of benefits to parents, since school performance information orients parents' decisions with regard to school choice. For school leaders and administrators, the findings from evaluation should provide relevant indicators to

support school-level actions and decision-making. Internal evaluation benefits are seen as primarily for schools, teachers and pupils with the aim of improving learning, teaching, and management within the school.

21. *Schools are increasingly held accountable to multiple authorities,* such as education ministries, local councils, and also to the wider community of stakeholders. The situation in England typifies the tendency towards multi-accountability. Schools are accountable to the central level through OFSTED inspections, to their local authority and to a governing body which includes representatives of key stakeholders such as parents, staff and representatives of the local community. Local authorities and the Secretary of State have powers of intervention if the school's performance becomes a cause for concern. There is also a market-driven accountability system, in which parents are provided with information on school performance (on-line OFSTED reports, and schools' test results, etc.) to inform their choice of a school. The Netherlands has followed this trend towards multi-accountability. Schools are accountable to central government for both budgetary matters and pupil achievement (the ministry having the right to restrict funding if school performance is poor). 'Organising bodies' are also involved in school evaluation and schools are accountable to families since freedom to choose a school is accompanied by a system of information provision for families (inspection reports and attainment test results are available on-line). Potential tensions and inconsistency of objectives may stem from these competing forms of accountability.

2.2 Responsibilities for school evaluation

2.2.1 Agencies with responsibility for school evaluation

Agencies responsible for defining criteria and standards

Standardisation of the criteria for school external evaluation in the context of school autonomy

22. There is a current trend in most OECD countries to base external evaluation on objective considerations, by providing evaluators with lists of criteria set by the central education authorities. When predetermined lists of criteria are established for external evaluation, they are either defined by chief inspectors (as in the Netherlands, Portugal, the United Kingdom, the Czech Republic and the Slovak Republic), or by departments within ministries or education authorities (as in the Flemish Community of Belgium, Catalonia, Andalucía and the Canary Islands in Spain, and Iceland).

23. Since the late 1990s, the criteria used for the evaluation of schools have been subject to standardisation in many of the countries in which the central educational authorities are responsible for evaluating schools. The majority of OECD countries thus report that they are developing national goals or standards of student performance (OECD, 2008c). This trend towards standardisation seems to equate to reinforcement of school accountability and increased professionalization of the evaluation service (Eurydice, 2007). This applies to Portugal since 1999 and to some of the Spanish Autonomous Communities (Andalucía, Catalonia and Castilla-La-Mancha). In Poland a uniform list of criteria has also been used since 2004. In England, the Inspection Framework drawn up by OFSTED sets out the requirements for evaluating schools and pupils' attainment and lists the criteria which form the basis for all school inspections.

24. In some countries, local education authorities have kept a pivotal role in the evaluation process but, since the mid 1990s, national standards and student attainment standardized tests have been developed as part of a national structure for school accountability. Paralleling the movement toward developing curriculum standards for students, standards for schools have also been developed for the purpose of school evaluation. This applies to Denmark, Belgium, Hungary, Canada, and the United States. For instance, Hungary rounds off the inspections carried out by local authorities with a new compulsory system of evaluating schools through standardised tests.

Schools have some leeway when it comes to defining criteria for internal evaluation

25. Whether internal evaluation is compulsory or recommended, there are generally instructions and recommendations regarding its implementation. Though schools are seldom compelled to use a predetermined list of criteria in the course of internal evaluations, a general set of aspects are usually defined by the education authorities, on the basis of which schools have to define their criteria for internal evaluation. However, schools typically have some leeway when it comes to defining criteria. Self-evaluation usually centers on the question whether schools meet their own objectives. When the results of internal evaluation are destined to be used by individual schools in order to enhance their quality, their staff usually plays an active part in defining the criteria. In a limited number of countries, internal evaluation is used either to report to the educational providers and education authorities, or taken into account in the course of external evaluation. In this case, the criteria tend to be partly determined by the education authorities, in order to make comparisons easier.

Agencies responsible for evaluating schools

Division of responsibilities for the evaluation of administrative and educational tasks

26. In most countries, school evaluation covers two main aspects: educational tasks and administrative tasks such as resource management (OECD, 2008b). Typically these two aspects are assessed by different agencies. In Poland, for instance, the inspectorate (*kuratorium* at regional level) evaluates educational tasks, while the *gminy* (municipalities), which are responsible for the management of compulsory education, deal with the evaluation of administrative tasks. Similarly, in Hungary, regional pedagogical institutes or independent experts authorised by the local authorities are responsible for evaluating educational tasks, whereas other experts, also at the request of the local authority, analyse the efficiency with which schools manage their budget. In lower secondary education in France, the school plan is evaluated by an inspectorate. School heads and, via them, all school activities are evaluated by the *académie* authorities, whereas evaluation of the management of capital and operational resources is carried out by regional auditors.

27. *Local authorities may also intervene when they are the school educational providers.* Since their role as evaluators is not always officially specified at central level, they may evaluate schools in accordance with their own interests. In Denmark, the municipalities evaluate aspects concerned with both educational and management tasks for which schools are responsible, whereas specialised bodies evaluate specific areas. In the partner country Estonia, the state supervisory agency is involved in those aspects of school activity concerned with teaching/learning, whereas the owners of schools (local authorities) are concerned with administrative aspects, the monitoring of compliance with legal standards and the management of resources. In Slovakia, the inspectorate is responsible for evaluating educational tasks, whereas each school founder monitors the financial aspects of schools. The school authorities, which are state bodies present in the regions and districts, delegate some of their responsibilities for financial auditing and the monitoring of organisational aspects to specialist departments.

28. *In some countries, responsibilities for school evaluation are shared.* In the United Kingdom for instance (England and Wales), Ofsted (in England) and Estyn (in Wales) regularly inspect the entire work of each publicly-funded school at least every six years. These inspections have three main purposes: to hold schools accountable to parents and the local community, to help schools plan for improvement and to provide information on the national state of education. Local education authorities (LEAs), on the other hand, have a legal duty to promote high standards of education, and a wide range of other responsibilities which, if they are to be exercised effectively, require them to monitor their schools. Though specific evaluation procedures are not prescribed, LEAs are expected to visit all their schools at least once a year to

discuss school improvement and target-setting as well as to provide early identification of schools causing concern.

Division of responsibilities for the evaluation of educational tasks between regional and central levels

29. In the majority of OECD countries, the central educational authorities are responsible for evaluating schools through the inspectorates. Inspection systems may be centralised or devolved. In New Zealand, Germany and Spain, schools are evaluated through inspections attached to the higher educational authorities – respectively the government, the Länder and the Autonomous Communities. In some countries, the inspectorate has a high degree of autonomy (in the Netherlands and the United Kingdom), although it remains ultimately accountable to the Ministry of Education (in the Netherlands and Scotland) or to the Parliament (England, Wales). Inspections can be coupled with local authority evaluations, as in the Czech Republic and the United Kingdom (England).

30. *In a few countries, the body responsible for evaluation is accountable to a decentralized authority.* This applies for instance to France, Austria or Poland. The inspectorate, which is run by the top-level education authority, also operates offices at regional level, for instance at the level of the *académie* in France. In Austria, though the inspectorate is a department of the Ministry of Education, Science and Culture, inspectors are affiliated to the federal school authorities in the provinces (Bezirksschulräte, district school board), Landesschulräte (province school board) and work within the jurisdiction of the *Landesschulrat* responsible for organizing school inspections in each of the nine *Länder.* In Poland, the *kuratorzy* are responsible for the evaluation of schools in each province *(voivodship)*. They implement the policy of the Ministry of Education, but report to the head of the province who reports to the prime minister.

31. Some countries share responsibility for evaluating schools or the educational tasks of schools between various levels of authority. This applies for instance to the United Kingdom and Hungary. This responsibility is shared between the municipalities and the national agency responsible for external evaluation in Sweden and Denmark (NAE and EVA respectively) and, in Iceland, between the ministry and the municipalities.

2.2.2 Other actors involved in the conception and implementation of school evaluation

Other actors involved in the implementation of external evaluation

32. Teams of inspectors occasionally call upon experts to support them in the course of external evaluation. When teams of evaluators in Germany, Austria, Scotland, and the Czech Republic need specialist assistance they can call upon experts in the fields concerned to join them on an *ad hoc* basis. Experts are concerned only with the specific issue they have been asked to investigate. In Austria, school inspectors may call upon other inspectors, as well as experts and teachers with special knowledge to support them. In Scotland and Northern Ireland, inspection teams also frequently include at least one "Associate Assessor" (practicing teachers, head teachers, or managers serving in other educational establishments). Their involvement adds an element of peer review to the inspection process. Teams of inspectors occasionally call upon staff who have not specialised in the field of education. Evaluation teams in the United Kingdom (except those from the LEAs) have to include a "lay member" in order to ensure that stakeholder views are fully taken into account. Lay members are representatives of the public, and as such are not involved professionally in education. The lay member can deal with such areas as relations with parents or with the pupil's comfort and their personal and social development.

Other actors involved in the conception of internal evaluation

33. Parallel to external school evaluation, many countries have recently developed internal evaluation, which can be defined as a type of evaluation where the professionals (teachers, head teachers) take the responsibility for the evaluation of their own organization (either their classes, or the school as a whole). At the school level, internal evaluation can be performed by a teacher or a group of teachers, by other members of the school's professional personnel, by the principal or other school administrators, or by a special staff member designated to serve as an evaluator. The teaching staff may carry out the school evaluation themselves or make use of external advisors. Educational umbrella organizations, national agencies, educational authorities and schools themselves have developed tools and methodologies to enable schools to evaluate educational quality autonomously.

2.2.3 External evaluators: qualifications and training

2.2.3.1 Background

34. An external evaluation of a school can be performed by the school district, a ministry of education, using professional evaluators, regional inspectors, or a district/state/national evaluation department. External evaluation can also be conducted by an independent evaluation consultant (Hungary) or evaluation firm, commissioned by the school itself or its governing board (Nevo, 2002). Requirements for school inspections and the criteria for appointment as a school inspector may operate from the central administrative level or from lower levels, such as regional offices or municipalities.

35. The predominant function of external evaluations conducted by inspectors is accountability. As inspectors provide feedback to schools about their strengths and weaknesses, inspections are also a vehicle for schools to improve. According to R. Standaert (2001), two main paths have developed historically. The first is exemplified by the United Kingdom and is characterized by a focus on the institutional quality of the school as a whole. The second path, observable in France, is characterized by a search for quality focused on teachers. In addition to these two primary approaches (i.e. full inspection of the school on the one hand, and inspection of teachers and principals on the other hand), other tasks have been given to the inspectorate, such as thematic inspection, inspection of branches of study, the supervision of test and examinations, monitoring of in-service training, or the development and maintenance of a database related on the education system.

2.2.3.2 Required qualifications and experience for candidates

36. *In most countries the candidates for a function as inspector are required to have experience in education or teaching*. Candidates are usually required to have a teaching qualification for the level of education which they are going to evaluate. In Germany, a member of staff responsible for supervision in primary, lower or upper secondary education (*Schulaufsichtsbeamter*) is required to provide evidence of the same qualifications as teachers at the school level concerned and must have completed several years of teaching service. In countries in which teaching qualifications are required, candidates must also have completed several years of service as teachers: for instance, eight years of experience in teaching are required in Slovakia, ten years in the French community of Belgium and in Greece.

37. When a central national inspectorate does not exist, inspections are carried out under the supervision of education providers at the local level. In some cases, it is up to local educational providers to determine their own criteria for the appointment of inspectors (Denmark, Hungary*)*. In Hungary, inspection of education is performed by independent experts of public education listed on the national list of experts (Országos Szakértői Jegyzék). Entry on the list must be applied for, and is conditional upon complying

with some basic criteria, namely teaching qualifications and a minimum of ten years of experience in teaching and education.

38. In countries in which teaching qualifications are not formally required, a relevant degree or professional experience in either training or advising is generally required. In England, a relevant degree and/or professional qualification within the relevant setting (teaching, training, inspecting or advising) are required in order to undertake inspection for the Ofsted. Though there are no specific requirements in the Netherlands, candidates with experience in the sector in which they are to work are preferred.

39. *Experience in school management or specialisation in evaluation may be required.* As a general rule, inspectors must have acquired several years of experience as head teachers or in a senior position in a teacher training institution in Germany. In Poland, the requirements for appointment as an inspector are a minimum of six years of employment in a teaching post and the completion of in-service training courses related to administration or management, or a minimum of 2 years experience in a managerial position at school. Similarly, all inspection teams from the Czech School Inspectorate must include one controller, namely an inspector specialised in resource management and administration. Controllers have to have a degree in law or economics and at least five years of professional experience in education. In New Zealand, although no teaching qualification is formally required to join the Education Review Office (ERO), Review Officers should have acquired previous experience in management.

40. Although inspectors are usually required to assess both educational and administrative tasks, no country requires formal qualifications in management or business administration. This apparent paradox may be explained by the following observation from the European Commission Education DG "where it is clearly stated that applicants must testify to their administrative qualifications, this is always in terms of their professional experience. It is expected that managerial competence should have been acquired in the course of previous work as a school head, (…) or in other posts in educational administration" (Eurydice, 2004).

41. In addition to experience in education or management, social or specific skills are recommended to be eligible for the position of external evaluator. Analytical skills are broadly required. Most countries place great emphasis on the knowledge of legislation. In multi-linguistic countries, candidates must be able to speak more than one language to facilitate contacts with other cultural communities. In Spain, inspectors have to accredit competence in the co-official language of the corresponding Autonomous Community. In the Czech Republic, inspectors are required basic computer literacy.

2.2.3.3. Recruitment process

42. In the vast majority of countries, external evaluators are permanent civil servants employed by the central (or regional) authority responsible for carrying out external evaluation. In the countries offering civil servant status to evaluators, candidates have to meet demanding criteria: they are generally expected to hold a degree, to complete a specialist course, to succeed a competitive examination and to complete a probationary period. Countries that entrust external evaluation to independent experts seem to accept candidates with a broader range of qualifications as long as they submit evidence of the expertise required in educational or managerial issues. This applies for instance to Denmark, the United Kingdom (for lay inspectors and registered inspectors), Iceland and Hungary. In England and Wales, there is a system of contracting out school inspections. Ofsted and Estyn are responsible for the management of the system of school inspection, under which school inspections are carried out by independent inspectors, engaged by inspection providers. These are commercial organizations or local education authorities (LEAs), who bid for contracts to carry out inspections. Their work is monitored by Her Majesty's Inspectors (HMI), permanent employees of Ofsted and Estyn.

43. *Inspectors are usually recruited on the basis of a competitive examination*. An interview and discussion with an examination committee is also a standard procedure in the recruitment. In a number of countries, a teaching test has been introduced for the recruitment. In France and in the French and Flemish Communities of Belgium, a written test is organised as a first selection, followed by an oral examination. In Spain, the competitive examination phase comprises three different parts which assess candidates' knowledge of teaching techniques, administrative issues and legislation on education as well as the command of appropriate techniques to carry out inspections.

44. In several countries, the entrance to the inspectorate is being widened by allowing non-teachers or school managers to apply. This is the case in Portugal, England, the Netherlands and for controllers in the Czech Republic. In Sweden, inspectors are recruited by the municipalities mainly among teachers and teacher trainers but can also be practitioners in other fields such as government and local authorities. In Scotland, inspectors from backgrounds in social work, youth justice, health and care have recently joined HMIE to deliver children's services inspections.

2.2.3.4 Initial training and in-service training for the inspectorate

45. Special training in evaluation is usually offered following recruitment or appointment. In France (in the case of the regional educational inspectors, IPR) candidates who pass the exam are assigned to an *académie* and undergo two years of alternate work and training at the national training centre. On the approval of the *recteur* at the end of the training program, they are certified as IPRs. In Portugal, in Ireland and in the French Community of Belgium, training of inspectors includes special courses related to evaluation as well as a mentoring process, involving guidance from experienced staff. Training courses include aspects related to administration or management in Spain, Portugal and the United Kingdom. In countries in which candidates have to undergo specialist training in evaluation prior to appointment, new inspectors are closely observed.

46. Systematic in-service training has been developed in a few countries, especially for the inspectors' first years of service. In England, Ofsted requires all appointed inspectors to complete a minimum of five days of professional development every year and keep up to date with inspection practice. In Slovakia, school inspectors of the State School Inspectorate (SSI) must complete in-service training. As a general rule, though a limited number of countries have a definite approach for the initial training of beginning inspectors little attention is paid to systematic on-the-job training (Standaert, 2001).

2.3 Evaluation procedures

2.3.1 Reference standards and aspects assessed

2.3.1.1. Standards

Educational standards for the external evaluation of schools

47. The term educational standard can be used in different ways (Döbert *et al*., 2004). Educators and policy makers usually refer to educational standards as student learning outcomes (knowledge, skills, disposition) they expect the education system should induce in the students at different grade levels. More specifically, evaluators consider standards as the desirable or acceptable levels (cut scores) of learning or performance that would inform whether or not the performance of a school is good enough. Standards formulate measurable expectations towards schools and instruction and therefore make schools more accountable for their achievement. Grading scales and required standards for schools and pupils might focus on national average or value-added indicators which take into account schools' background and pupils intake. Educational standards might also take the form of educational goals for specified grade levels.

Performance standards define the achievement on a proficiency scale, and finally provide assessment standards. Evaluators may also focus on content standards, related to competency dimensions.

Standards for the internal evaluation of schools

48. *Internal evaluation should usually relate to the school's own aims, as set out in its educational plan*. In most countries, the general objective of the internal evaluation is to report goal attainment and measures planned for increasing goal attainment. Self-evaluation is therefore strongly connected to the school development programme. The obligation to consider how far the educational policy of the school has been implemented is consistent with one of the aims underlying internal evaluation, namely supporting schools in critically appraising and developing their own quality. It is also fully consistent with the need to balance schools' empowerment with accountability measures in the context of school autonomy.

49. Schools may be recommended to rely on the standards and criteria defined by external evaluators when they carry out internal evaluation. As a general rule, regard must be paid to documents specified at a level higher than the school: these documents may be specific concerns, indicators (France), or the criteria used by the inspectorate for external evaluation (Catalonia, England). In France, in the case of evaluation carried out by the school head, the only restriction in determining the focus of evaluation is that due regard must be paid to a set of centrally formulated concerns. The Netherlands, the United Kingdom, Slovakia, which have centrally determined lists of criteria for external evaluation, recommend that schools use them in their internal evaluation. Similarly, in Sweden, internal evaluators have to refer to general guidelines for quality reporting laid down by the Swedish National Agency for Education. This is consistent with the fact that in several countries, such as the Netherlands, external evaluation partly relies on the judgement formed during internal evaluation.

2.3.1.2 Aspects assessed in the external evaluation of schools

50. In many countries, the definition of aspects to be assessed in the course of school evaluation follows Stufflebeam's CIPP evaluation model (context, input, process and products). "Context" is usually related to readiness to learn, public attitudes to education, and the role of school in the community. Inputs include facilities and equipment, educational choice, financing, teaching characteristics and classroom characteristics. Processes are related to expectations and attitudes, learning measures, teaching arrangements, parental and community involvement. Finally, products include student academic performance and post-schooling outcomes.

Processes assessed in the external evaluation of schools

51. *A wide range of school processes can be assessed.* Certain aspects such as teaching and learning processes, or guidance and support for pupils, are virtually always part of the evaluation grids predetermined at the national level. The parameters may include equity concerns, such as the treatment of pupils with special needs. The educational approach of the school is often considered, through the examination of the school plan or the school annual report. The management of human, financial and material resources is also often taken into account. Other aspects such as the functioning of bodies with mixed representation or internal evaluation are typically less relevant (Eurydice, 2004). The classroom practice of teachers is evaluated in accordance with parameters that relate to the content of their lessons and the teaching methods and materials employed by them, the quality of their interaction with pupils, respect for their personalities and the development of their self-confidence.

52. *Some aspects covered go beyond the educational responsibilities of schools:* they may relate to the relations with parents in order to estimate if schools seek to associate them with decision-making or take account of their opinions. School relations with other schools, local institutions or the local community in

general may be considered too. Parameters concerned with 'leadership', such as the coordination of teaching activity, the participation of teachers or pupils in decision-making, the communication of information, the pursuit of common objectives and the sharing of a 'school ethos', aim at evaluating the social and professional atmosphere within the school. In a minority of countries, such as England or Wales the improvements carried out since the previous inspection are also evaluated.

Outcomes assessed in the course of external evaluation

53. *Quantitative outcome parameters* relate to the academic performance of pupils (results in tests and examinations), their rate of absenteeism, the drop-out rate, and indicators such as the success rate in progressing from one year to the next, or the proportion of pupils who qualify for special education. Pupils' performance in national tests (*e.g.* Scotland) and the results of exams organised by schools can be used (*e.g.* the Flemish Community of Belgium, the Netherlands, Scotland and Slovakia) in order to assess the performance of the school. Assessments enabling the comparison of students and schools might encompass performance in several fields: assessments may measure specific aspects of student performance and quality outcomes; be curriculum based on specific subjects; or be cross curricular problem solving type. When inspectors take into account the results obtained by pupils in specific subjects, they focus on the language of instruction and mathematics in most cases.

54. *Qualitative outcome parameters* are associated with the cognitive development of pupils or their social behaviour. Qualitative outcome parameters relate for instance to the theoretical, communication or analytical skills acquired by pupils while at school. The cognitive skills of pupils are qualitatively evaluated as outcomes of educational activity in the Flemish Community of Belgium, in Slovakia and in the Czech Republic, usually by observing pupils directly in the classroom. The social skills of pupils are evaluated as an outcome of education by observing their attitudes (Slovakia, the Czech Republic), their involvement in extra-curricular activities in Spain (the Canary Islands), or their relationships in the United Kingdom. In the Flemish Community of Belgium, Spain, the Netherlands, the United Kingdom and Slovakia, schools are evaluated in accordance with both quantitative and qualitative outcome parameters. In Scotland, the inspectors evaluate schools by means of a composite performance indicator which integrates several types of outcome parameters. These parameters include the performance of pupils in national examinations, the quality of their learning, and the progress they achieve. In England and Wales, under the Education Act (2005), Her Majesty's Chief Inspectors are required to take into account, among other things, the educational standards achieved by schools, the spiritual, moral, social and cultural development of pupils, and the contribution made by schools to the well-being of pupils.

2.3.1.3 Aspects assessed in the course of internal evaluations

55. *As with external evaluations, internal evaluations may focus on processes, outcomes or both*. Such evaluation is usually concerned with both educational and administrative tasks. In some countries the aspects to be addressed are only marginally externally dictated as with the quality of the educational process in Slovakia or the quality of management in Greece. Where more external direction is given, aspects to be evaluated typically include the obligation to take account of the performance of pupils and of the relations of the school with the school community or parents (Eurydice, 2004). For example, in Austria, internal evaluation should cover teaching and learning, the classroom and the school environment, the partnerships of schools, school management, the professionalism of staff and staff development. Indicators expressing input, processes and results must also be prepared. In Spain, where self-evaluation is compulsory for all educational institutions since 1990, each Autonomous Community proposes the contents on which self-evaluation must focus. In some cases, they refer to the whole of the institution; in others, they deal with partial aspects of the school activity, such as curricular competences, the equity and efficiency of education, school coexistence, etc.

56. *Some countries emphasise the importance of results by encouraging schools to use performance indicators* (France, Portugal, Iceland and the United Kingdom). In France, a set of indicators for the guidance of secondary schools enables them to measure their specific characteristics and to compare themselves with other schools in the country. In England, schools are also expected to analyse pupil performance using benchmarks and performance indicators. In addition, as regards improvement, schools are expected to set specific targets for pupil performance, evaluate the quality of teaching against national criteria, plan objectives and targets for improvement and evaluate their effectiveness (Ofsted, 1998).

2.3.2 Evaluation instruments

57. A wide range of instruments are available to external evaluators. The observation of lessons is typical in those systems based on Inspectorates. The method of observation is generally not formalised. Interviews are also usual in all inspection bodies. This pertains to interviews with school managers and individual or groups of teachers. Parents are interviewed in England, the Flemish Community of Belgium, the Netherlands, Scotland, and the Czech Republic. In Ireland parents are consulted as representatives of the Boards of Management. In some systems, such as England, Scotland, the Czech Republic, the Netherlands and the Flemish Community of Belgium, pupils or representatives of pupils are also interviewed.

58. *Systematic questionnaires* are used in the majority of the countries. School managers fill in a questionnaire during the preparation phase in England, the Flemish Community of Belgium, the Netherlands, Portugal and in Scotland. In Scotland school managers and teachers also fill in a questionnaire after the inspection in which they can give their opinion of the inspection process (Standaert, 2001). In several countries, parents and pupils may reply to questionnaires when data are gathered. This type of involvement is particularly used for internal evaluations. Parents are given a questionnaire in which they are asked about their view of the school in the Czech Republic, England, Portugal and Scotland. In the Netherlands, the United Kingdom and Norway, parents and pupils may be consulted when data for internal evaluation are collected. In Norway, they may also be members of groups set up specifically to carry out internal evaluation. In Scotland, parents are consulted about priorities regarding quality improvement.

59. *Using test data for inspection* is also usual. In some countries, where some forms of external examinations exist, inspectorates use these results to confront schools with averages, benchmarks, or targets. In some countries, such as Ireland, the inspectorate does not use the results of the national examinations as evidence in full inspections or for purposes of ranking schools. These results are only used for national surveys. During inspections, the inspectors may apply their own tests; or check the results of school-based pupil testing. In Scotland for instance, the inspectorate has its own tests for primary education and education up to 14 years. For secondary education, the results of central testing are used to compare schools' results and even as a means for self-evaluation. Similarly, England has a sophisticated set of central and external tests for different stages in the education system. The inspectorate in England uses these data for the ranking of schools.

60. *Self-assessment*: In several countries, the external evaluation of a school is based partly on the results of internal self-evaluations. As a general rule, the results of self-evaluation are used by external evaluators for a preliminary appraisal of a school before it is visited. For instance, external evaluation in the partner country Slovenia draws on the report of the school head, which analyses the achievements of the school in the previous year and sometimes evaluates its performance. Similarly, external evaluation undertaken by the local authorities may be partly based on evaluation conducted by schools. In Denmark, the majority of municipalities ask schools to prepare an annual report that they use in their own evaluation of schools. The aim of internal evaluation in Sweden is to establish whether schools have achieved national educational objectives and local objectives as well as their own aims, and to make proposals for improving their performance with respect to those objectives. This entails preparation of a *quality report* which, along with other data, provides a basis for evaluation of the school concerned by its municipality. As regards self-

evaluations, monitoring progress towards objectives in school requires a wide variety of techniques. Among those in use are review teams, questionnaires, checklists, interviews, team teaching, classroom observation, shadowing pupils as they work in various subjects, systematic planning and reporting and discussion groups.

61. In recent years there has been a debate in many countries concerning the best approach to evaluating schools. The various methods being tried range from external control and inspection to self-evaluation and internal regulation. In practice, most education systems are moving towards a combination of these methods, involving a degree of external monitoring of internal self-evaluation mechanisms (McNamara, O'Hara, 2005).

2.3.3 Methodology and procedures

2.3.3.1 Stages of evaluation

62. *Inspections typically consist of several stages*: notice of inspection to schools, pre-inspection and background documentation, site inspection visit and observation, reporting and follow-up (OECD, 2008b). In a Eurydice study it is assumed that the processes of evaluation include four stages, namely, the gathering of relevant information, the assessment, the drafting of the evaluation report and the implementation of changes (Eurydice, 2004).

63. *As a general rule, approaches to evaluation include a preliminary stage of investigation*. In many countries, evaluators systematically collect various documents that they examine before they actually visit schools. Qualitative and quantitative information is usually gathered through different sources of information. In almost all countries, evaluators consult documents made available to them by schools. These documents may be plans relating to the educational or organisational policies of schools (the school plan, curricular plan, information and communication technology development plan, etc.). They may also be factual data concerning schools (results of internal tests or examinations, data on class composition, or the financial management report). Documents produced by schools for the benefit of outsiders (brochures and the annual report of the school to the parents) may also be consulted. Finally, evaluators may consider the findings from evaluations or audits conducted by schools. Evaluators also examine documents and reports emanating from the central authorities. They may consist of complaints about a school or a member of its staff, or statistical or financial reports. The results obtained by pupils in national tests and examinations, as well as previous evaluation reports (e.g. Poland, Ireland), are also a source of information for evaluators in most countries. Finally, evaluators in some systems (e.g. the United Kingdom, the Czech Republic, Slovakia, or the Canary Islands in Spain) send questionnaires to persons who belong to the school community, including the school head, the management team, teachers, parents, pupils or members of school boards or committees. In England, for instance, the process gathers a wealth of detail on all aspects of school life including enrolments, pupil background, the range of classes offered, the results of national tests and exams, and the results of interviews with parents and school governors.

64. *The second stage is the site visit*. As a general rule, the school head and the teachers visited are previously informed about an intended school inspection. In England, for instance, most of the time of the Ofsted inspectors is directed at reviewing classroom teaching and pupil reaction. Beyond examining the teaching and educational standards attained, the inspection team must review and comment on school leadership, the management of school resources and finance and the spiritual, moral, social and cultural development of pupils (Rosentahl, 2004). In the course of inspection, evaluators have different ways of organizing their work. Several evaluating bodies have formed teams of inspectors in order to broaden the range of skills to be mobilised and to evaluate schools more thoroughly. Inspectors who act as a team are coordinated by one of them, usually an experienced inspector. The composition and size of inspection teams are determined by the head of the inspectorate. In the Netherlands (secondary schools) and the

United Kingdom (England and Wales), the number of team members may be as high as 10 and 15 respectively (Eurydice, 2004).

65. Another stage in the evaluation procedure in some countries involves a discussion between evaluators and certain school members on the findings of the evaluation. This discussion occurs before the final evaluation report is written and gives schools an opportunity to react to it or clarifying certain points. The possibility offered to certain members of a school to comment on the findings of the evaluation may be regarded as an indication that the latter is conducted in a spirit of cooperation with the entity evaluated (Eurydice, 2004). The category or qualifications of persons able to consult the preliminary report or attend the meeting at which its findings are made known orally may be precisely regulated. In all countries, the school head at the very least is consulted. The management team and the school board may also be involved in the discussion or may consult the final draft version of the report. Parents are seldom involved in the discussion. For instance, in the Netherlands, the management team may be accompanied by up to ten other members of the school during the meeting at which evaluation findings are made known to schools in the Netherlands. Only the management team is allowed to comment the preliminary report.

66. *In some countries, another stage in the procedure involves follow-up to the evaluation*. The follow-up to the evaluation may be regarded as a means available to the educational authorities enabling them to carry out their responsibility for monitoring and improving the quality of schools. In countries in which the follow-up to evaluation is concerned solely with schools experiencing major problems, the procedure would appear as a means to ensure that all schools achieve a minimum standard level of quality. In countries in which follow-up occurs in all cases, it may be regarded as a means of supporting each school, irrespective of its circumstances, in its efforts to improve quality (Eurydice, 2004). The procedure of following up an evaluation is distinct from the secondary evaluations which generally occur shortly after the (first) evaluation and are primarily concerned with confirming or invalidating its results. Their purpose is therefore not to evaluate the progress achieved by schools but rather to check that the conclusions of the evaluation were appropriate. This type of evaluation generally occurs whenever serious shortcomings have been identified during the first visit and the school concerned runs the risk of disciplinary action. This applies to the Flemish Community of Belgium and the United Kingdom (England and Wales). During the follow-up stage, the same evaluators or colleagues who belong to the same evaluating body examine how far schools have achieved the objectives they have been set during their evaluation, or check that they have complied with the recommendations made to them. In a minority of countries, systematic follow-up occurs after each evaluation (Poland, Scotland, and in some Autonomous Communities in Spain, such as Catalonia and Andalucía).

67. *The evaluation process concludes with the reporting stage*. The final report may be publicly disclosed over the internet. In 2008, results of evaluations undertaken by school inspectorates were published by 12 OECD countries, but only in Iceland were they used by the government to publish comparisons of the performance of individual schools (OECD, 2008a). In the Flemish Community of Belgium, the Czech Republic, England, Ireland, Korea, the Netherlands, New Zealand, Portugal, Scotland and Sweden and the partner country Estonia the results of evaluations undertaken by school inspectorates were published, but were not used for the creation of comparative tables of school performance.

2.3.3.2 Character and status of evaluation

68. *Types of schools covered*: Public and private sectors are usually submitted to the same evaluation criteria. The character and periodicity of school evaluation usually differ depending on whether primary or secondary schools are being considered.

69. Two-thirds of OECD countries have regulations that require lower secondary schools to be inspected regularly, and slighter fewer countries require schools to conduct periodic internal evaluations. To evaluate

school performance, about half of OECD countries have both of these regulatory requirements. Two-thirds of OECD countries organize periodic standardized assessments of students in order to evaluate student performance. National examinations for lower secondary school students are also organised in half of the OECD countries (OECD, 2008c). In the majority of OECD countries, schools are regularly inspected: this is the case in Australia, the Flemish Community of Belgium, the Czech Republic, United Kingdom, France, Germany, Greece, Iceland, Ireland, Korea, Luxembourg, Mexico, the Netherlands, Portugal, Spain, Sweden, Switzerland, and Turkey. Schools conduct regular self-evaluations in Australia, Denmark, United Kingdom, France, Germany, Greece, Hungary, Iceland, Ireland, Japan, Korea, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain and Sweden (OECD, 2008c).

70. *Few countries have no systematic school evaluation.* In Italy for instance, schools are not accountable to a specific body (except for the control of administrative and accounting regularity that is carried out by the Boards of Auditors) but are strongly encouraged to perform self-evaluations. A standardised evaluation system is currently being developed under the auspices of the National Institute for the Evaluation of Education, Training and Teaching (INVALSI), in order to provide national benchmarks for schools.

71. *Self-assessment is mandatory in an increasing number of countries*: this applies to the French and German Communities of Belgium, France, the Netherlands, Spain, Greece, Sweden, Norway, Poland, Czech Republic, Slovakia, Portugal, Hungary and in the partner country Estonia (OECD, 2008a). In other countries, self-evaluation is voluntary (Austria) or strongly recommended (United Kingdom, Italy). The status of internal evaluation may also be variable depending on the local authority concerned (Australia, Canada, Denmark, and Finland). In Australia, most states require schools to report annually on performance. In Canada, internal evaluation status is variable from province to province.

2.3.3.3 Periodicity of evaluation

72. As a general rule, the frequency of the internal evaluation is not regulated by law. Regular internal evaluation is strongly recommended, since internal evaluation must be carried out in a long-term perspective. In Iceland, the external evaluation of internal evaluation methods used by compulsory and upper secondary schools is organised by the Ministry of Education, Science and Culture at five-year intervals. Whenever school self-evaluations are required, these are generally required annually, unlike evaluations by school inspectorates which tend to be required only every three years or so. In Japan, there are no requirements for the frequency of school evaluations but these evaluations still take place in a substantial proportion of schools (OECD, 2008a).

2.4 Using evaluation results

2.4.1 Using evaluation results for accountability

2.4.1.1 Consequences of evaluation results for schools

73. The results from evaluation can be used to support accountability (including rewards and sanctions) and market-mechanisms (through the publication of school results and parental choice) in the allocation of resources.

74. *The consequences of external evaluations for schools vary widely across countries:* a school may be given informal recommendations (*e.g.* Ireland, Iceland), lose its recognition or financing (*e.g.* the Flemish Community of Belgium, the Czech Republic), or be given the label of a "failing school", *i.e.* a school requiring special measures. The consequences may also involve the possibility of school shut down or entail financial sanctions or rewards for schools or individual members of the school staff. The impact of performance and accountability measures depends on the objectives and context in which they are developed. In some co0untries, such as in the Netherlands and the United Kingdom, accountability is

sought in the form of strengthening consumer control (league tables, school guides) more than in the form of active measures (rewards and sanctions) from higher administrative levels (Scheerens *et al*., 1999).

75. *Extra-support can be targeted at needed schools*. School evaluations were considered to have had a significant influence upon assistance provided to teachers to improve their teaching skills in Australia, the Czech Republic, England, Mexico and Turkey (OECD, 2008a).

76. *Disciplinary action can be taken against schools when performance is judged to be unsatisfactory.* The sanctions associated with school accountability systems can imply the possibility of school shut down, loss of status, or financial sanctions for schools or individual members of the school staff. In the Czech Republic, for instance, if gross deficiencies or failures are ascertained, the Chief School Inspector may submit a proposal for the exclusion of the school concerned from the state school sector. In this case, schools receive no further public funding and are no longer authorised to award official certificates. Similarly, in the Netherlands, if the Inspectorate identifies serious shortcomings, it submits an inspection report to the Minister on the school concerned, accompanied by recommendations as to measures to be taken. The Minister may decide to take administrative action, including penalties, such as a funding cut. In the United States penalties for repeated failure to meet performance standards have become established in all states as a result of the No Child Left Behind Act of 2001 (NCLB): schools are subject to sanctions if any one subgroup of its students fails to meet the state-determined targets for adequate yearly progress (AYP) targets. If schools fail to make AYP for two consecutive years, their districts must identify them for improvement and give their students the opportunity to transfer to other schools. The law further mandates corrective actions for schools that fail to make AYP for a third consecutive year (US Department of Education, 2002). Beyond NCLB-mandated penalties, two thirds of state education systems have their own, additional policies to penalize persistently low-performing schools (Chiang, 2009). More generally, school evaluations were considered to have had a high influence upon: school financing in the Flemish Community of Belgium and Sweden; the provision of financial rewards or sanctions to schools in the Flemish Community Belgium; the remuneration and bonuses received by teachers in the Czech Republic; and on the likelihood of school closure in the Flemish Community of Belgium, the Czech Republic and England (OECD, 2008a).

77. Despite the emphasis on educational outcomes, the allocation of public funds to schools according to the results they achieve is not a widespread practice. School evaluation and student performance measures are mainly used to provide performance feedback to schools and educational authorities. As a general rule, they have little influence on school financing and other financial implications such as changes to the school budget, provision of financial rewards or sanctions for schools, or remunerations received by teachers. Only a few countries including Korea and the United States, reported using accountability information to provide financial sanctions to schools (OECD, 2008a). In the United States, the development of assessment-based accountability programs was one of the key features of education reform in the last decade. Several states and a few local districts have introduced school-based performance incentive programs. Such programs rank schools based on their success in increasing the achievement of their students, provide financial rewards for the schools at the top of the ranking, and may impose sanctions on the schools at the bottom of the ranking. Over 30 states offer sanctions or rewards to schools based on their students' scores on standardized exams (Sims, 2005).

78. *Assessing programs designed to enhance equity:* Programs specifically designed to enhance equity in education are the subject of specific evaluations which may lead to sanctions or changes in the school status. In France, schools are given *éducation prioritaire* status on the basis of the socio-economic characteristics of students and the learning outcomes. An evaluation (Bénabou *et al*., 2004) found that the *zones d'éducation prioritaire* (priority education areas) had not had a significant effect on school outcomes. The "Ambition réussite" reforms adopted in 2006 by the Ministry of Education aim at directing extra resources efficiently on the basis of evaluation results. Schools will be evaluated more systematically and

may obtain or lose priority education status depending on evaluation results. In the Flemish Community of Belgium, disadvantaged schools are provided with additional resources too. The Equal Educational Opportunities (GOK) Decree of 2002 aims at countering exclusion and social segregation. Schools with a high number of disadvantaged pupils may receive additional support and funding for a period of three consecutive years, if at least 10% (in primary and lower secondary education) or 25% (in upper secondary schools) of the school's pupil population is made up of target-group pupils who meet equal education opportunities indicators. Self-evaluation by the schools forms an integral part of the three-year GOK cycle, and the achievement of goals by schools receiving extra resources is evaluated every three years by an independent body. Schools may lose the additional resources if evaluated negatively.

2.4.1.2 Publication of the findings from school evaluation and its consequences for school accountability

79. The findings of external school evaluations are made public or made available upon request in several countries. Pressure on schools is stronger when individual school reports are made public. Information may be published by educational authorities, inspectorates, individual schools or the media. In Portugal and in the Netherlands, pressure exerted by the media has played an important part in the publication of school evaluation results. The results of student assessments by school are also published in Australia, the Flemish Community of Belgium, England, Italy, Korea, Scotland and Turkey and in partner country Slovenia (OECD, 2008a).

80. The information can be publicised among the entire population (for example, via the press or internet), or merely circulated among the parents of pupils. Reports are usually published on the Ministry's or on the Inspectorate's website. This applies to Sweden, the United Kingdom, Denmark, and the Czech Republic. They may also be directly sent to stakeholders. In England, the governing body of the school must send a copy of the inspection report to all parents within five working days of receiving it. The report is then published on Ofsted's website. In Iceland, it is up to the municipality and the school to decide whether and how they present the results of external evaluation to stakeholders (parents, pupils, the community, the school board, etc.). However, results are available on the Ministry's website too.

81. *Information may be accumulated specifically for purposes of comparison*, or be of interest solely to the schools concerned. In France, the assessment of schools has evolved with the annual publication of performance indicators for *lycées* (upper secondary schools). In England and Turkey the results of student assessments were used by the government for the creation of comparative tables of school performance (OECD, 2008a). In addition to individual school inspection reports, HMCI publishes an *Annual report* based on inspection evidence on *Standards and quality in Education in England* which receives a great deal of media coverage. Each year since 1993 the HMCI report has publicly identified particularly successful schools ("outstanding providers") that have shown outstanding performance in inspections during the year. These schools are highlighted because they "have received an outstanding inspection report and have continued to perform very well in tests and examinations given the circumstances of the schools"(OFSTED, 2003). These reports aim at providing incentive for improvement through comparison. In addition to the accountability motive (accounting for performance to the tax payer, the government and parents), publication of school performance results also provides information to parents which can be used as a basis for informed school choice. School improvement is often supposed to be promoted by the competition between schools resulting from the publication of school performance indicators, in combination with open enrollment (Visscher, 2000).

2.4.2 Using evaluation results for improvement

82. *Follow-up to external evaluation and its consequences for school improvement:* In addition to the accountability motive, results from school evaluation can also be used as tools to reveal best practice and identify shortcomings in order to encourage schools to improve. Conceptually speaking, different kinds of

evaluation orientations may be distinguished with respect to school improvement (West, Hopkins, 1997). On the one hand, evaluation *of* school improvement may focus on the outcomes of improvement efforts or on the fidelity of process implementation. The school evaluation has a summative approach. On the other hand, evaluation *for* school improvement is used during the process of school improvement in order to further shape this process. The orientation is formative rather than summative and implies follow-up and support provision.

83. In a minority of countries, the school being evaluated alone is responsible for following up the evaluation and decides whether improvements should be introduced. This applies to countries in which efforts are made to strengthen the centrality of school self-assessment (Austria, Denmark). External evaluation leads to the formulation of recommendations or instructions. In some countries, schools are under no obligation to follow them. In Austria, inspection is advisory in nature: school inspectors look into the quality of teaching and the implementation of administrative tasks. The main outcome of school inspection is the counseling of teachers and of the headmaster. A verbatim report is to be drafted, if necessary, on the contents of counseling, especially if serious shortcomings were detected. However, schools are not obliged to implement measures geared to improving their practices, which is consistent with general efforts to strengthen self-evaluation skills in Austria. In Iceland, the evaluators do not insist that a plan for improvement should be drafted, following the evaluation of internal evaluation methods. It is up to the municipality to decide whether and how improvements should be introduced.

84. In most countries, evaluators check whether schools respond appropriately to the recommendations or instructions for improvement issued following the evaluation. In the Flemish Community of Belgium, when a school has been fully inspected, a follow-up inspection takes place after three years. The follow-up is based on an adjustment plan the school has drawn up.

85. *The extent of follow-through activity depends on how effective the school was deemed to be and its capacity to improve.* This applies for instance to New Zealand, the United Kingdom or the Netherlands. In England, good and outstanding schools are subject to a lighter touch inspection, whereas schools with the most severe weaknesses (including low standards of pupil achievement and weak leadership and management) are described as needing special measures. Their progress is monitored and they are re-inspected one year after being placed in special measures. In Scotland, HM Inspectors will spend most time monitoring and supporting those schools causing most concern. In other cases they may ask the local authority to take the lead in monitoring and reporting on progress in addressing needs identified by the inspection. HM Inspectors may sometimes revisit a school to explore further and disseminate very good practice. In the Netherlands, inspection is intensified at institutions with problems. "Proportional supervision" is conceptualized in the Dutch Supervision Act of 2002. The frequency and form of school inspections are based on the quality of the school and the risks of a decline in quality: poorly performing schools are inspected sooner and more often than the schools that perform better (Janssens, van Amelsvoort, 2008).

86. In some countries, schools are asked to draft a plan for improvement when performance is judged to be unsatisfactory. Using the findings of the evaluation, schools draft a plan setting out the measures that should be adopted to improve their situation. However, there is no scope for disciplinary action against schools, except in the case of the United Kingdom. In Denmark, the school being evaluated by the Evaluation Institute is obliged to draw up a follow-up plan which must relate to the recommendations for quality improvements in the evaluation report. In Spain, the plan for improvement is drafted by the school board in Catalonia, the management team in Andalucía, and by the special evaluation committee in the Canary Islands (Eurydice, 2004). It is drafted by the school and local authorities in Scotland.

2.4.3 Developing competencies to use evaluation results effectively

87. *Preparation of schools*: According to the empirical literature investigating the relationship between school leadership and student achievement (Hallinger, Heck, 1998), leaders' practices can have direct effects on school outcomes as well as an indirect influence through the way leadership has an impact on school organization and school culture. School leadership is needed to carry out internal evaluations, as well as to develop competencies within the school to take the most of evaluation results. For instance, school leadership is needed to engage the teaching staff in the use of data and analytical skills. Technical assistance, through the provision of benchmarks and guidelines, and training for school leaders are being developed, chiefly in those countries committed to promoting school self-assessment.

Development of competencies to carry out internal evaluations

88. In most OECD countries, internal evaluation is carried out by the school head, generally assisted by the school staff. As a teaching background does not necessarily prepare for evaluation practice, principals might lack knowledge in personnel and financial management and the skills required to carry out self-assessment and use evaluation results, such as setting goals and measuring progress (OECD, 2008c). Similarly, the school team is composed of teachers and other educators whose professional training is not in evaluation and for whom evaluation represents only a small part of their job. The major weakness of such a team is in its lack of evaluation skills and experience (Nevo, 1998). In addition to technical support, school leaders have to be provided with training in the use of the instruments and in the management of the evaluation process. Consequently, an increasing number of countries provide leaders with support and training to prepare them for internal evaluation as well as to help them to use evaluation results effectively.

89. *Initial training for internal evaluators:* Training in school self-evaluation is not a common practice. In most OECD countries, heads of schools have to undertake special initial training to prepare them for their responsibilities. This provision may include training in evaluation, but this is not a widespread practice. Although internal evaluation is not compulsory in the United Kingdom, in Scotland an applicant for a head teacher post must hold the Scottish Qualification for Headship, which includes training in evaluation as well as other aspects of education management.

90. *In service training courses are provided for head teachers and teachers in several countries.* Certain *Länder* in Germany, some Autonomous Communities in Spain, and some municipalities in Iceland, Norway and Hungary offer courses in evaluation to those interested. Regular in-service training seminars for evaluators are offered in Greece. Although internal evaluation is not compulsory in the Flemish Community of Belgium, in-service training is offered to secondary school heads. In Austria, the *Pädagogischen Institute* (in-service training institutions) offer courses in evaluation to all teachers and headteachers. Training in internal evaluation and school management is compulsory for principals. Similarly, in Portugal, courses in evaluation are provided by school association training centres and higher education institutions (Eurydice, 2004).

Support available to schools

91. In addition to training, supporting measures are available to schools to carry out and use evaluation results. Guidelines for internal evaluation are usually made available either by the national inspectorate or the Ministry of education. In some countries, schools are supplied with statistical indicators intended to monitor the education system, which allows comparisons of the performance of their pupils with that of other pupils. School heads may also be provided with performance indicators (France, Portugal, and the United Kingdom) supplemented with references to national or local averages so that they can compare the performance of their school with that of other schools in the same area or in similar context.. In Scotland, in addition to statistical indicators, the Standard Tables and Charts (STACs) provide comparative

information on attendance at school, as well as school costs per pupil and school leaver destinations. Some countries make information available to schools on the Internet in order to assist them with their internal evaluation. This applies to Sweden, the United Kingdom, Austria, Norway and the Czech Republic. In Austria, for instance, the *Qualität in Schulen* Internet Platform supplies schools with information and tools for both evaluation and data analysis, and provides a forum for the presentation of results.

92. Schools are usually allowed to call upon the assistance of external players, such as academic experts qualified in evaluation, educational advisers or consultants, teacher trainers, or teachers from other schools. External players may be employed by schools to undertake the entire internal evaluation, or merely provide schools with advice as regards the methodology to be used to undertake self-assessment. As a general rule, schools voluntarily rely on assistance of this kind and bear the financial burden of these extra resources. However, some countries place human resources at the disposal of schools. In Finland and Scotland, the local authorities may pay experts or consultants employed as resource persons for internal evaluation (Eurydice, 2004). The experts placed at the disposal of schools may also come from national institutions specialized in education or evaluation. This applies to the Federal Institute for Educational Research, Innovation and Development of the Educational Sector (BIFIE) in Austria. There may also be the opportunity to benefit from the temporary appointment of individuals working in teacher education institutions, management teams of other schools, or local authority officials. This is the case in Germany, Finland, Sweden, the United Kingdom and Iceland. In England and Wales, for instance, LEAs are expected to provide support and advice for schools causing concern. Elsewhere, the extra human resources might be teachers from other schools. In Norway and in two Canadian provinces (the Province of Prince Edward Island and Nova Scotia), peer review is used in many schools as a method of internal evaluation. Similarly, in the Netherlands, Finland and Sweden, schools may call upon the teachers or heads of other schools for assistance with their internal evaluation.

3. ARTICULATION BETWEEN SCHOOL EVALUATION AND OTHER TYPES OF EVALUATION

3.1 Use of school evaluation results for the evaluation of the school system

3.1.1. Use made of external school evaluation results

93. The results of external school evaluations usually lead to the production of a national report destined to the central education authorities for the purpose of evaluating the system. In most countries, inspectors rely on the results of the school evaluations they undertake to draw up public reports on the education sector as a whole. Such annual reports are prepared in the Czech Republic, England, the Flemish Community of Belgium, Portugal, the Netherlands, Spain, the Slovak Republic and France. In New Zealand, the Education Review Office (ERO) publishes the information gathered in all school reviews over a period of six months. In Scotland, Her Majesty's Inspectorate of Education (HMIE) publishes evaluations of the school system as a whole every three years. The 2006 "Improving Scottish Education" report uses school evaluation results in order to promote improvement across the system. This widespread practice illustrates the use made of school evaluations in view to assessing the school system. In Iceland, though the independent experts conducting the two approaches to evaluation (school evaluations and external evaluation of internal evaluation procedures) do not themselves produce national reports, the Icelandic Ministry of Education nevertheless uses their findings to monitor the school system.

94. When school evaluations are carried out by decentralised educational authorities, evaluations do not systematically lead to the production of a national report with the purpose of assessing the school system. In some countries, such as the United Kingdom (England and Wales) and Hungary, the evaluation conducted by the local authorities is used primarily by these authorities or the schools themselves. In England, Her Majesty's Chief Inspector (HMCI) makes an annual report before Parliament on the outcomes of inspection by Ofsted. In Wales, Estyn (Her Majesty's Inspectorate for Education and Training in Wales) provides advice based on inspection evidence, to inform the Welsh Assembly Government on educational progress.

95. In some countries, findings from external school evaluations performed at the local or regional level are an important source of information for central education authorities. This applies to Denmark, Sweden and Poland. In Denmark and Sweden, schools are externally evaluated by the municipalities. The findings from these evaluations are then processed by a national agency specialized in education, the Danish Evaluation Institute (EVA) and the Swedish National Agency for Education (Skolverket). These findings are then used by the central education authorities as a source of data on the state of the school system. In Poland, the *kuratoria* responsible for evaluating educational tasks at the regional level drafts a report every five years on the quality of education in the region. These reports are sent to the Ministry. At the same time, information gathered by the *gmina* at the local level via evaluation of administrative and management tasks are used by national authorities for various monitoring activities.

96. In a minority of countries, findings from the external evaluation of schools are not used by the central education authority to monitor the school system. This applies to those countries in which no central school inspectorate has been established. In Hungary, external evaluation consists of a report which serves to

prepare a decision by the school maintainer. In lack of a national inspectorate, the national body responsible for evaluating the school system (the Centre for Educational Evaluation and Examination) chiefly relies on pupils' achievement. In Austria, the school inspectorate primarily works at the level of the administrative districts and the provinces; its main outcome is the counseling of the teaching staff. The Federal Ministry for Education, the Arts and Culture operates no central school inspectorate, though a new institution was founded in 2006 to take account of the increased evaluation requirements, the Federal Institute for Educational Research, Innovation and Development of the Educational Sector (BIFIE).

3.1.2. Use made of findings from internal evaluations

97. The results of school internal evaluation are seldom used by education authorities or specific national bodies for the purpose of evaluating the school system. The central education authorities directly use the results of internal evaluations for monitoring purposes in a minority of countries. The findings from internal school evaluations are used as documents for an appraisal of the school system in the French Community of Belgium, the Spanish Autonomous Community of Andalucía and Iceland.

98. Activity reports provided by individual schools might be examined by central authorities, but are not used to evaluate schools or the entire school system. Findings from internal evaluations may have an indirect impact on the monitoring of the school system. Local government or the educational providers, or both, in certain countries, may use the results of internal evaluation to draft reports for education authorities at a higher level. The municipalities in Sweden and Norway use findings from internal evaluations to inform central education authorities. In Sweden evaluation reports drafted on the basis of the school activity report prepared by the school head are sent to the NAE. In Norway, most municipalities use the reports prepared by schools when communicating with the central Ministry.

3.2 Articulation between school evaluation and teacher evaluation

3.2.1 The individual evaluation of teachers and the collective responsibility for school teaching quality

99. Teacher evaluation may be interlinked with school arrangements to improve teaching quality in a range of ways. First, teacher evaluation can serve the direct purpose of improving the teaching process within the school when teacher evaluation is carried out as part of school internal evaluation, and distinct from the formal individual assessment of teachers. Teacher evaluation for improvement purposes as part of internal school evaluation is usually carried out by the school head or the school management team. It may also be carried out by peers or experienced teachers: "peer reviews" can be conducted as a way to expand the pool of evaluators and use evaluators with a background in teachers' subject and grade level. Though teacher evaluation for improvement is chiefly designed to enhance classroom practice, regular feedbacks by the principal and experienced peers might help both teachers and schools to identify priorities for both teacher and school improvement. Results from this kind of teachers' assessment can be used to identify teaching needs and contribute to the definition of the school plan in order to improve the teaching process within the school. This applies for instance to Poland, where judgments are formed of teachers as part of the internal evaluation of schools as entities.

100. Second, school evaluation and teacher evaluation are interlinked when school performance results are used to hold individual teachers accountable for their work. Whether as part of internal or external evaluation, pupil results in national standardised assessments are becoming a basis for judging the performance of a school and, indirectly, its teaching team, as in Sweden or Scotland. Policies for the individual appraisal of teachers based on pupils' results have also emerged. For instance, in Hungary, the results achieved by pupils in standard examinations may lead to extra bonuses for their teachers since 2007. The current shift from processes to results-based teachers' evaluation is a step towards closing the gap between individual and collective forms of accountability. In an increasing number of countries,

monitoring mechanisms are concerned with teachers both as individuals and members of the school team (Eurydice 2008). In Chile, for instance, teachers are rewarded collectively when they work in schools which are identified as high-performing by the National Performance Evaluation System of Subsidized Schools (SNED) (OECD, 2005).

101. Third, school arrangements and teacher evaluation are interlinked when the contribution of the individual teacher to school development is assessed as part of the individual assessment of teachers. In a number of countries, teachers are increasingly assessed both as individuals personally responsible for their students and as members of the school teaching staff. In Chile, since August 2003, all teachers in schools belonging to the municipal system are evaluated via the Teaching Performance Evaluation System (*Evaluación del Desempeño Profesional Docente*) on the basis of criteria which include teachers' professional responsibilities.

102. Fourth, school arrangements and teacher evaluation can also be linked when professional development activities for teachers which result from their individual appraisal are planned in the context of the general development plan of the school. Several countries now link professional development to the developmental priorities of the school and co-ordinate in-service education in the school accordingly (OECD, 2005).

3.2.2 The evaluation of teaching as part of schools' evaluation

103. Though improving school outcomes depends, among other things, on improving the quality of classroom teaching, teachers' evaluation is seldom integrated into the evaluation of schools as entities. Whether and how individual teacher appraisal should be integrated into the evaluation of schools as entities is open to discussion (Eurydice, 2004).

104. Evaluation (whether internal or external) targeted at schools sometimes provides an opportunity for evaluators to evaluate the personal work of teachers and then provide them with personalised feedback. When teachers are assessed on an individual basis during the evaluation of schools, the evaluation is generally external. This is the case in Germany, the United Kingdom, Ireland and the Czech Republic. Teachers' evaluation is usually based on conversations with teachers and visits to classes. In this context, teachers' evaluation may focus on teachers' subject knowledge, classroom management, or teaching skills.

105. *The observation of lessons is typical in most inspection systems,* though the emphasis is different depending on whether they serve as an integral school inspection or as an individual assessment. In school inspections, inspectors look for a large sample of teachers and generally use a more or less standardized observation list. However, in countries such as Sweden and Iceland, observing teachers in the classroom is not a method used by external evaluators for gathering the information needed to evaluate educational tasks within the school. Interviews are widespread and generally involve school managers and individual or groups of teachers. After observing the work of teachers, external evaluators generally provide them with personalised feedback, besides submitting their overall report on the performance of schools. In Poland, assessments are formed of teachers as part of the internal evaluation of schools.

106. When individual teachers are assessed in the context of school evaluations, the assessment itself is generally normative from the point of view of the school overall effectiveness but not that of individual teachers. In the Czech Republic, Denmark, England, the Flemish Community of Belgium, the Netherlands and Scotland, teacher assessments can take place in the context of school evaluations, but are then solely used to form a judgment on the quality of the school.

3.2.3 Use made of teacher evaluation results for school evaluation

107. Results of individual teacher appraisals are seldom used for external school evaluation. This may be explained by the fact that in most countries, teacher evaluation and school evaluation serve distinct purposes. Individual teacher evaluation usually has a limited focus as it is designed to identify priorities for individual professional development. The use of results from individual teacher evaluations is more spread in the context of internal school evaluations. In some countries, such results are taken into account by school heads when assessing teaching quality in their school. Teacher assessments can thus assist with the development of improvement plans for schools.

4. ISSUES IN DESIGNING AND IMPLEMENTING SCHOOL EVALUATION APPROACHES

108. This section discusses several contentious issues in designing and implementing school evaluation. The more controversial aspects in school evaluation schemes are related to whether or not student outcomes should be used as a measure of school performance, whether the findings from school evaluation should be disclosed to all stakeholders or not, and the kind of accountability systems that should be used in order not to create undesired effects on school improvement. The advantages and drawbacks of different approaches are related to the purpose of school evaluation emphasised. Finally, the implementation difficulties that may result from different stakeholders' views and interests are discussed.

4.1 Advantages and disadvantages of approaches to school evaluation

4.1.1External school evaluation

109. *Allowing comparison*: The external evaluation of schools has typically the advantage of granting the possibility to compare performance across schools and to assess performance against reference standards. However, external evaluation is often criticized for its narrow scope and its tendency to focus on commonalities rather than uniqueness in its attempt to seek comparability and generalization (Nevo, 2002). Such interpretation might overlook the local perspective, reflecting the special needs of the school. External evaluation is limited when it comes to interpreting specific findings regarding local perspectives, context, needs and constraints. By contrast, the strength of internal school evaluation teams is in its educational and pedagogical background, its familiarity with the educational and social context of the school, and its non-threatening relationship with the school administration (Nevo, 1998).

110. *Producing incentives:* The threat of sanctions on low-performing public schools has become a dominant form of incentive by which policymakers seek to raise student achievement. Insofar as repeated failure is penalized, schools face a powerful incentive to raise their performance ratings. The mechanisms schools have at their disposal for improving their ratings that are consistent with policymakers' intent are those that reform the inputs and processes of educational production within failing schools, but schools may choose to manipulate the accountability system in ways that raise school or student performance scores without contributing to students' knowledge and skills (Chiang, 2009)

111. *Pressure and stress:* Anecdotic evidence from teachers' and school leaders' interviews suggests that the evaluation process is stressful and demanding. In France, for instance, according to the National Council for School Assessment (Döbert, 2004), the current system is considered as causing "suffering and distress to both those being assessed and the assessors". The educational staff of the school being assessed may feel personally called into question and often do not receive either recognition for their work or the help they could expect from an evaluation.

112. *Direct and opportunity costs:* The costs of evaluation are hardly trivial. These costs include direct costs of implementation (financial and material) and opportunity costs related to time and resources allocation. Costs obviously vary widely depending on the kind of evaluation schemes. Comprehensive classroom evaluation systems are time-consuming, more demanding on teachers under evaluation and more expensive than once-a-year principal evaluations or evaluations based only on student test scores. In

England, for instance, the gross budgetary allocation to Ofsted was £150.4m for 1997–98. It may be argued that the time of school managers and their teams is precious and best used for pedagogical tasks rather than collecting information and developing indicators (Döbert *et al*., 2004).

4.1.2 Use of student results to evaluate schools

113. *Data on pupil performance can be used as a basis for school comparison,* and provide benchmarks for individual schools. According to the proponents of standardized tests in education, centrally defined tests are objective and reliable enough to provide criteria for valid comparisons among individual students. National examinations can serve as a crude indicator of school performance, but provide no indication of the value-added of a school, unless they are properly adjusted. The level of average test scores or pass rates may also be used as a measure of school success. Because test scores and socio-economic status are highly correlated, however, such an approach risks to measure in part the socioeconomic characteristics of the students in the school rather than the contribution of the school to student learning. A different approach focuses on the rate of each school's improvement during the year (as measured by changes in the test scores of, say, third graders one year to third graders the following year) relative to a school-specific target rate of improvement. This is the case for instance in Kentucky in the United States. The main drawback of this approach is that it fails to account for differences in the mix of students from year to year. By contrast, value added measures focus on changes in the performance of students from one year to the next (and hence require annual testing of students). The value-added approach explicitly recognizes that students who enter a grade with below-average achievement may leave the grade with below-average achievement, even if school administrators and teachers have made effective use of their resources and the student has made significant gains during the year. Value-added measures are generally appealing to schools and teachers, because they only hold the school accountable for students that are enrolled in the school for at least a year.

114. *Such indicators are limited in terms of how validly they reflect the quality of schools*: Although assessments enabling the comparison of students and schools might encompass performance in several fields, it may be argued that reliance on standardized test and examinations fails to capture the full spectrum of outcomes of the school. Student results are generally expressed by means of quantitative data, which can only reflect a limited number of school quality aspects since not all school results can be measured quantitatively.

115. Achievement testing systems serve a number of purposes. For policy-makers they provide a general picture of student achievement in the country, identifies the most underachieving schools for special assistance programs, evaluates the effectiveness of such programs, and serves as a basis for rewarding high-achieving schools. Information from student assessments may be used by the educational authorities to improve the allocation of human and financial resources and facilitate the targeting of resources to schools that are in trouble. As regards improvement, student results may serve diagnostic and formative purposes for school leaders and teachers and be used to shape improvements in school management and classroom teaching and learning. Student results are also a way to empower schools, since leaders might use them to support decisions about how to improve practices and how to direct resources. They can be used to assess school objectives, adequacy of curriculum, and adequacy of teacher performance. Test-based accountability systems are also supposed to strengthen incentives for teachers to commit themselves to helping all students to meet important centrally defined standards and fulfil goals within the national curriculum. As regards accountability, information from student results is likely to raise public awareness and prompt citizens to bring pressure to bear on ineffective schools (Stetcher, 2002). For parents, achievement testing systems are intended as an accountability tool that strengthens involvement in their children's education and that orients their decisions with regard to school choice. Publication of the results of student performance on external tests and examinations also adds to the stakes and pressure for improvement. When targets are defined for specific groups on the basis of their social background or

minority status, like in the United States, test-based accountability systems are likely to enhance equity within the school.

116. *Tensions between the accountability goals and the use of student results to improve school outcomes*. The uses of student results - for accountability and for improvement - may sit together uneasily. The sanctions associated with test-based accountability systems ─ which imply the possibility of school shut down, loss of status, or financial sanctions for schools or individual members of the school staff ─ may stifle innovation for school improvement or produce unintended side effects. For instance, school focused accountability systems can lead to strategic responses on part of teachers and schools. These responses might lead to the reallocation of efforts among topics or skills to meet the test standards at the expense of improvement (Koretz, 2005). More particularly, "high-stakes" testing are associated with ambiguous effects on teachers' practices (Jacob, Levitt, 2003). Focus on high-stakes tests may introduce incentives for preemptively retaining students and increasing special education placements of low-performing students in special programs which are outside the accountability system. Certain practices can reduce the meaningfulness of test scores as indicators of students' performance and lead to "score inflation". Indeed, targeted teaching to those skills that are represented on a test can raise scores without increasing students' mastery of the broader domain (Stetcher, 2002). Focus on high-stakes tests may lead to a narrowing of the curriculum. If the same test form is used repeatedly, teachers may become familiar with the specific items that appear on that form and shift their instruction accordingly ("teaching to the test"). A potential reaction to national or international testing is to revise curricula to enhance greater match between test content and curriculum. An example of this effect of testing is when Ontario is believed to have taken action to revise its curriculum to match the TIMSS test after 1995 TIMSS administration (Döbert, 2004).

4.1.3 Performance-based funding

117. *Competition and market-type forms of accountability*. Schools may be put under increasing pressures for performance and competition to attract students. Such an outcome can be considered as an advantage, provided that market competition is assumed to improve technical and allocative efficiency in the use of resources. Arguments in favor of performance-based funding may be found in England, at the time when the policy of Local Management of schools (LMS) introduced in 1990 led to devolution of management and budgetary powers to schools, while there was also a strong trend towards increase in central control through increased evaluation and accountability arrangements. In this context, the Conservative Government claimed that increased competition and accountability would improve attainment because of the greater incentive the system provided to managers and teachers compared to the previous system of administrative allocation of resources. Such a structure supports the principle that successful schools will attract more students, producing a linkage between the resources delegated to the school and the educational output through the mechanism of parental choice.

118. *Impact on school staff motivation*: Just like other accountability oriented reforms, performance-based funding may shift teachers' motivational goals : teachers' intrinsic motivation to teach may be partly displaced by such extrinsic goals as money and "winning" the school's ranking competition when the stakes are high (Leithwood *et al*., 2000).

4.1.4 Public release of results and school "rankings"

119. *Publication of the results of school evaluations is a matter for debate:* Whether the results from individual school evaluations should be used privately by teaching staff and educational administrators or should be published remains controversial (Rowe 2005, Visscher, 2001). School accountability to families developed from the middle of the 1990s with the publication of results in a growing number of countries. Publication can take different forms: on-line inspection reports, reports of local authorities, and publication of standardised student test results by school. The consequences of the public release of results depend on

the nature of the information published and the way in which the information is published. For instance, unadjusted raw school performance scores may have a stronger negative impact than value added performance data.

120. *Positive effects are associated with the public release of school results.* By introducing accountability, public posting of evaluation and tests results tend to promote parental and government pressure for quality (Haegeland *et al*, 2004). It might drive learning by raising the stakes for schools, pupils and teachers if evaluation reports and test results are used to rank schools publicly. Information disclosure is expected to make teachers and administrators increase their efforts to improve performance. Competition to attract pupils may drive improvements in school quality where market mechanisms enable parents to choose a school on a fully informed basis. In addition to influencing motivation and behaviour at the school level, disclosure of information about school performance is expected to rectify asymmetric information about school performance – and thus increase efficiency in resource allocation on the basis of this information.

Disadvantages associated with the public release of school and student results

121. The main criticism on the publication of school results refers to their limitation in terms of how validly they reflect the quality of schools (Visscher, 2000). League tables must be interpreted with caution in association with context data in order to avoid stigmatisation. When important regional differences or differences between the public and private school sectors exist in terms of regulations, resources, staff recruitment and pupil admission, comparisons between different types of schools may be unfair, unless all relevant factors are taken into account. Indeed, part of student outputs are influenced more by factors extrinsic to the schools than by those for which schools might be held accountable. Though value-added information is a pre-condition for valid comparisons between schools, it might not be sufficient for knowing the true performance of schools since full adjustment for all relevant factors is impossible (Visscher, 2000).

122. A second kind of criticism on the publication of school results refers to their limitation in terms of usability. Usability issues relate to the extent to which schools and students results can be used by various stakeholders for the purpose of school improvement and accountability. As regards school choice, accessibility to public school performance information is not equally distributed across parents: certain groups of parents remain unaware of their existence, or lack the skills to properly interpret it. This is a serious drawback if positive effects of the release of this information are expected. Schools may also differ in the possibilities they have at their disposal to improve educational provision. Since publication of crude league tables and school labelling are not conducive to the formulation of within school improvement strategies, external pressure can be damaging for schools that do not have the skills to initiate changes, especially if they do not receive adequate support. Educational authorities should be concerned that potential users are properly informed of the shortcomings of output 'league tables', even adjusted ones. An over interpretation of a set of rankings where there are large uncertainty intervals, can lead both to unfairness and unwarranted conclusions about changes in ranks (Visscher, 2001). There is also doubt as to whether school performance publications stimulate schools to improve their functioning. Such indicators are summative by nature and usually do not point schools to the cause of and remedy for underperformance. "League tables'' and the ''shaming and naming'' of failing schools puts all the emphasis on sanctions, which seems to run counter to what is known on the factors that motivate individuals (Rutter, Maughan, 2002).

123. *Misinterpretation of school results and rankings may be damaging for equity*. Publication of tests and evaluation results by means of "league tables" can damage equity and distort information. The approach emphasizes competition among schools, rather than providing a focus on needs at a population level. This is likely to discourage schools from accepting or retaining children with special needs and

sidesteps both the desirability of considering what is needed to bring about change and the time frame required for change to be evident in the indices of success being used in league tables (Rutter, Maughan, 2002). The effectiveness of schools that serve predominantly low-income or at-risk communities is masked by the effect of family background. Working in a school that has been classified as low-performing can have negative effects upon school staff and students and impede school improvement. Increased competition between schools to attract students may have social consequences. Well-off parents shunning low-performing schools may reinforce social polarization into good and bad schools, increasing inequity. Market mechanisms might lead to overcrowding in some schools and increased segregation in areas where schools are low-performing.

124. The publication of student and evaluation results touches on the tension between the accountability function and the improvement function. External pressure resulting from public release of information may lead schools to use performance evaluations for improvement. However, just like the use of students' results, external pressure as a result of the publication of school results may promote strategic behaviour.

4.2 Difficulties in implementing school evaluation

4.2.1 Conflicting perceptions of school evaluation

125. Different stakeholders may have different expectations from schools and support the use of different criteria for measuring school effectiveness (Scheerens and Bosker, 1997). Difficulties in implementing school evaluation might stem from conflicting perceptions of stakeholders.

4.2.1.1 Conflicting perceptions on external school evaluation

126. Teachers and school staff may argue that evaluation schemes constrain their autonomy. Teacher unions may oppose imposed evaluation schemes, especially when sanctions are at stake. In the United States, for instance, the No Child Left Behind Act has been controversial for a variety of reasons. Practitioners disapprove of the sanctions implied by NCLB. For instance, the largest education association (the National Education Association, NEA) is opposed to the use of financial penalties for low-performing schools, and rather suggest that hard-to-staff schools, especially those with high concentrations of disadvantaged students, should be provided supports and resources to help students succeed, including additional targeted funding to attract, retain and form quality teachers. Though few educators or policymakers find faults with NCLB'S goals of improving learning opportunities and outcomes for all children, some provisions of the law have been widely criticized, and several states filed suit in 2005 to be excluded from its provisions entirely. The most widely held criticism is that federal expectations for student performance have been set so high that virtually all US schools will fall short of NCLB requirements (Barrett Brian D., 2009). State and local policymakers also complain that meeting federal mandates will require enormous investments which are not coming from the federal government (Kochan, 2007).

127. Although the presumption of inspectorates is that the recommendations and findings of the inspections will lead to improvement of the quality of schools, for some critics, adverse results are to be expected. It may be argued that the stress and energy dissipated for only a few hours of observation and interaction is detrimental to both staff and pupils. In England, for instance, Ofsted argues that inspection processes have improved school procedures and management; provided incentives for schools to improve the delivery of secondary education; and encouraged poor school administrators and teachers to leave the profession. By contrast, Ofsted inspections have been reported to be "more likely to cause stress and strain, and a drop in effectiveness, than to lead to improvement" (Rosenthal, 2004). Similar opinions are expressed by teaching professionals in other countries. In France, the current system of assessment of

school quality is considered as causing "suffering and distress to both those being assessed and the assessors"(Döbert *et al*., 2004).

4.2.1.2 Conflicting views on use of student results to evaluate the school

128. Opponents of assessment-based accountability programs such as teacher unions argue that they lead to a narrowing of the curriculum, or teaching to the test, that decreases overall learning and is unfair to underprivileged students. Advocates of high stakes accountability tests argue that teaching to the test is appropriate if tests are properly constructed to measure achievement. They claim that a yardstick for student achievement provides teachers and administrators with incentives to help students learn. Teachers may also be opposed to evaluation based mostly on student test results. In the United States, where most states offer sanctions or rewards to schools based on their students' scores on standardized exams, the development of assessment-based accountability programs such as those included in the provisions of the 2001 No Child Left Behind Act has aroused opposition from teacher unions. Both the National Education Association (NEA) and the American Federation of Teachers (AFT) are strongly against using student test scores to evaluate schools and individual teachers. However, teacher unions are more tolerant of other schemes: in 2007 the AFT's largest affiliate, the United Federation of Teachers (UFT), which represents New York City's teachers, agreed to a two-year pilot program that awards performance bonuses to schools rather than individuals.

4.2.1.3 Conflicting views on the public release of results and school "rankings"

129. Public release of results arise political and ethical issues (Visscher, 2001). According to some experts, it is unethical to publish school quality data (either public performance tables based on raw scores or value added tables) because of the potential harm done by such labelling effects, as long as one cannot properly adjust for all relevant contextual factors in order to precisely indicate how well schools perform. Indeed, a more relevant issue is to what degree schools do or do not meet the standards considered important. On the other hand, it can be counter-argued that it would be paternalistic to collect school performance data and decide that only certain stakeholders (policy-makers, schools management team) can have access to it at the expense of others such as parents. Not all countries publish student performance results: in Germany, access to school final examination results has been reduced or prohibited by the ministries. In several countries, such as Denmark and Poland, publishing evaluation results, namely student performance results, is a matter for debate.

130. More generally, in all countries in which publishing evaluation results is a matter for debate, proponents claim that this enables parents to choose schools on an informed basis, while opponents to the practice of publishing evaluation findings, such as teacher unions, often use arguments based on a rejection to see education governed by market mechanisms. As regards teacher unions, the strategy of labeling some schools as "failing" or "good" is generally considered negative for the profession as a whole. In Canada for instance, although there is general public support for large-scale testing, teachers have vehemently opposed it mainly because of the high-stakes associated with it. In some provinces teachers resent using the test results to rank schools, which indirectly reflect on their teaching effectiveness. This resentment may have explained the teachers' negative attitude towards large-scale mandated testing (Döbert, 2004).

4.2.2 Implementation difficulties

131. Evaluation systems are intended to improve achievement through three mechanisms: strong incentives for school staff to improve their performance, high-quality information to guide these improvement efforts, and technical assistance to help schools that are unable to reach the goals on their own (Hamilton *et al*., 2005). In practice, these mechanisms might be lacking, which might compromise the implementation of recommendations for improvement.

4.2.2.1 Difficulty of combining the accountability and improvement functions

132. *External accountability mechanisms might emphasize compliance to procedural requirements instead of improvement*. Increased pressure stemming from evaluation schemes might inhibit commitment strategies and create incentives for mere compliance to administrative requirements at the expense of improvement and innovation. Negative effects of a control strategy on schools outcomes are reported by Leithwood (2002). Conceptually speaking, Rowan (1990) distinguishes two models of school organization, referred to as the control and commitment strategies. The control strategy involves "the development of an elaborate system of input, behavior, and output controls designed to regulate classroom teaching and standardize student opportunities for learning"; the expected result is an increase in student achievement. By contrast, the commitment strategy rejects bureaucratic controls as a mode of school improvement and instead seeks to develop innovative working arrangements that support teachers' decision making and increase teachers' engagement in the tasks of teaching.

133. Usability of evaluation results to make informed decisions and set agendas for action might be problematic, especially when it comes to transforming summative assessments into formative tools (Zimmerman, Dibenedetto, 2008). Although the purposes of accountability and improvement are complementary, feedback for improving education requires different kinds of information and specific capabilities than feedback for ensuring accountability. For assessments such as standardized tests to be useful, diagnostic information is needed.

134. *Data need to be provided in a timely manner*. Data from summative examinations administered at the end of students' tenure are likely to have little impact for the cohort just tested. Although results from national assessments that are delivered late in the school year may be useful for policy decisions related, for example, to resource allocation, they are of limited use to instructors (OECD, 2009). The time frame for reporting results of external assessments is often incompatible with the requirements of using diagnostic information to adapt instruction to student needs (Rowan, 2001), unless national assessments are conducted at the beginning of the school year and intended to serve diagnostic purposes like in France and the French Community of Belgium.

4.2.2.2 Resistance by some educational agents

135. *Resistance to the constraints on teachers' and school providers' autonomy:* Teachers and teacher unions can act as an obstacle on the development of evaluation schemes. Anecdotic evidence shows that teachers emphasise the constraints an evaluation scheme places on their autonomy, and are particularly anxious when it comes to evaluating their practice (Scheerens *et al*., 1999). Similarly, local authorities may oppose schemes imposed by central educational authorities (OECD, 2008c).

136. *Incentives for school staff might not be conducive to school improvement.* The failure of numerous reforms can be explained by inconsistencies between their objectives and the behaviour of the teachers and school heads that implement them. If reform design is inadequate, objectives are inconsistent with established programs, costs to individuals are excessive compared to the expected benefits, then school actors are unlikely to commit themselves to its implementation. Teacher support and acceptance of accountability systems and evaluation schemes is thus a challenge to implementation: teachers might support, passively accept or oppose a given evaluation approach. Imposed change is believed to create a "culture of compliance" (Datnow *et al*., 2000, Leithwood *et al*., 2000) among teachers at the expense of innovation. Teachers who are constrained in ways likely to reduce their own intrinsic motivation to teach may behave in more controlling ways and be less effective in teaching their students. By contrast, if teachers are involved in planning and implementing evaluation schemes, they are more likely to sustain reform efforts. Furthermore, perceptions of the teacher as intrinsically motivated increase the chances of students' being intrinsically motivated as well (Leithwood *et al*., 2000).

4.2.3 Lack of appropriate information and training to guide evaluation and improvement

137. *Implementation difficulties are likely to stem from insufficient information and training.* Evidence from a survey based on teachers' self-reports in three American states (California, Georgia, Pennsylvania) in 2004 suggest that most teachers are lacking appropriate support (Hamilton *et al*. 2005). The survey provides a snapshot of instructional practices in the context of the external accountability systems required by the NCLB Act. Teachers were asked about the usefulness of external progress tests: only about half of the teachers said they received training on how to use results for instructional decision making, and of those who did receive training, only about half found it useful. The results also suggest that some teachers have increased their use of "undesirable practices" such as more time teaching test-taking strategies.

138. Appropriate skills are required on the part of teachers and school leaders to use performance data and translate evaluation results into practice. Indicators are useful in showing trends and differences among education systems, but provide no information regarding reasons for the observed trends and suggestions for reversing the trends or reducing gaps. Testing does not immediately translate into good information for schools on the challenges they face, and even less into support for using data to guide school improvement (Townsend, 2007). Schools may lack capacity and training to understand and use data effectively to inform improvement. Scepticism towards data amongst educators based on mistrust or fear of data and evaluation is also likely to increase implementation difficulties (Campbell, Levin, 2008).

4.2.3.1 Lack of training for internal evaluators and leaders in charge of implementing recommendations for school improvement

139. Schools are more likely to develop capacity for improvement and accountability if they are in an environment that provides support for these factors. Most schools and local authorities lack the kinds of information systems that would provide high-quality data on student outcomes (Townsend, 2007). In addition, leaders in charge of implementing measures for school improvement may not receive adequate support everywhere. In decentralised systems, leadership development is the responsibility of local or regional authorities, which might make it more difficult to ensure consistency and equity across schools. Leaders may get different support depending on location. In Sweden, for example, school leaders working in different municipalities can have different opportunities to attend in-service training. In socially disadvantaged areas with less study tradition, challenges for schools might be larger and lead to under provision. A similar situation is reported in New Zealand, where schools have high levels of autonomy (OECD, 2008c).

4.2.3.2 Lack of support to assist schools in using evaluation results effectively

140. Effective feedback has to be provided in a systematic way to enhance achievement. Promoting the better use of existing systems to assess the school system at all levels (national, local) requires that the evaluators make efforts towards communicating properly their reports, tools and results, and that the actors strive to integrate results in processes to improve the system. In France, for instance, the current system of assessment of school quality is often criticized for its lack of effect and clarity. According to the National Council for School Assessment (Döbert *et al*. 2004), the inspection rates are dissimilar across regions, and the intervals between inspections can be long, particularly as a result of the insufficient number of inspectors. The conditions under which inspections are carried out can differ considerably, the assessment criteria are not always known to the schools receiving the inspection, and are defined at the local level, or even individually, by each inspector. Given the rate at which they take place, the results of the different assessments are put to very little use: reports are often received very late and are not the subject of any discussion.

141. *Schools might lack appropriate technical assistance.* For instance, in the United States, in the context of the federal No Child Left Behind Act, an increasing number of schools face the challenge of meeting higher performance standards without significant increases in funding or support (Bowen *et al*., 2006).

4.2.3.3 Legitimacy of the agents in charge of undertaking school evaluation and implementing changes

142. Lack of moral authority may have an important influence on the success of policy implementation. Evidence from a case study in the Canadian province of Ontario (Leithwood *et al*., 2000) makes clear that conflict and hostility between educators and the government about the nature of the accountability policies as well as the non consultative process used to both develop and introduce them is likely to introduce a pervasive bias into educators' judgments about the desirability of policy outcomes.

143. School heads may lack legitimacy, support and leadership capacity to implement measures conducive to school improvement, especially where principals are *primus inter pares* who return to the teaching profession once their service is finished. In countries where the position of principal is temporary, such as Spain and Portugal, there has not been much training neither for leadership in general, nor for evaluation in particular (OECD, 2008c).

4.2.3.4 Poor alignment of evaluation systems

144. Difficulties may stem from poor alignment of standards, curriculum and assessment, as well as misalignment of views on the role and purpose of evaluation. Most schools operate in multiple accountability systems. The different systems are seldom aligned, and might be in conflict with each other. Schools have to choose which ones to favour in any given instance (OECD, 2008c). Well-aligned systems enable evaluators to draw valid conclusions about student performance, or to adapt resources to better meet identified needs (Stecher *et al*., 2000). A well-aligned system implies comprehensive and clearly defined standards for student performance. If standards are poorly designed, teachers are more likely to focus on tests, thus narrowing the focus of teaching (OECD, 2009).

145. The complexity of school systems (including differences in regional and organizational contexts of schools) may be a barrier to alignment of standards, curriculum and assessment (Baker, 2004). Empirical evidence from Finland makes clear that where education systems are decentralized, local evaluation activities vary considerably between different municipalities. As regards Finland, local evaluation has been wide and versatile in the major cities, whereas evaluation has been rather limited and without adequate coordination in small and poor municipalities (Döbert *et al*., 2004). This problem is particularly acute regarding internal evaluation. Even though the dimensions and criteria for self-evaluation have been defined, their significance in practice is questionable. Although they are likely to serve as a development tool, self-evaluations, as such, have not yielded an adequate basis of reliable and valid data for educational indicators in Finland. Similarly, lack of state by state comparability was reported as one of National Assessment of Educational Progress' (NAEP) main shortcomings (OECD, 1999). Although diversity of education and indicator systems makes it inappropriate for comparisons, there is no agreement about how alignment of assessments and content standards should be evaluated (Linn, 2005).

4.2.4 Ways to overcome obstacles

4.2.4.1 Ways to overcome the potential perverse effects of the publication of results

146. Since publication of school-level results may pose risks to equity, one possible improvement to raw test scores is to assess the added value of the school in terms of results taking account of the characteristics of the student population. This approach is being pursued in Norway to evaluate the national education system: the National Education Database (NUDB) established in 2002 by Statistics Norway follows

individuals through the education system, and is connected to data such as family background. The database allows for analysis of student flows, drop-out, completion and how this is related to social background. Similarly, England is moving to indicators which adjust for pupil intake and school background (RAISEonline). Fairer comparisons are possible thanks to the provision of contextualized performance information related to schools with similar profiles. However, the use of such indicators has some drawbacks. Empirical evidence in two American states (North and South Carolina) suggests that value-added measures of school effectiveness exacerbate any existing incentives for teachers to avoid schools in which students are on average low performing in favour of those in which students perform better on average (Ladd, Walsh, 2002).

147. In order to optimize the publication and usage of school performance indicators, some experts recommend that a wide range of school process features and output characteristics should be published in order to take into account the many dimensions of school quality aspects. However, the publication of high quality performance indicators requires building and collecting large data banks, which imply that large sums of public money are being spent, whereas the effective advantages of publication are unclear (Visscher, 2001).

4.2.4.2 Providing more support for using evaluation results

148. *Supporting school staff in understanding and appropriating the evaluation*. Accountability systems are beneficial for school improvement providing that they engage the skills and commitment of the school staff. School leadership is needed to make external accountability beneficial for school improvement. This involves school leaders developing skills in interpreting data and test results in order to implement strategies for improvement. Self-evaluation and improvement are important responsibilities carried out by school leaders (OECD, 2008c). As regards internal evaluation, it seems important to provide training not only in the use of the instruments but also in the management of the evaluation process.

149. *Providing schools with benchmarks*: In several countries such as the United Kingdom or France, performance indicators are being developed in order to allow school leaders to implement the recommendations for improvement formulated in the course of school evaluation. These indicators are also designed to provide schools with tools to carry out self evaluations. Indeed, the construction of indicators and assessment tools requires technical skills that establishments and their managers do not always have at their disposal, while the time of these managers and their teams is best used for pedagogical tasks rather than collecting information and developing indicators. This is the reason why most of the indicators proposed take the form of a return of information from schools, which provide data for national information and management systems and receive automatically in return personalised indicators which are ready for use and accompanied by references (national, regional or departmental averages) which allow them to situate themselves against others (Döbert *et al*., 2004).

4.2.4.3 Overcoming the lack of legitimacy of the evaluators

150. *Clarifying the criteria used for school evaluation*: Anecdotic evidence shows that those evaluated may emphasise the lack of legitimacy of the evaluators. School staff may also complain about the lack of clarity of the criteria used, and what are perceived as arbitrary statements from the evaluators. In France, for instance, according to the National Council for School Assessment (Döbert *et al*., 2004), the system of assessment is not fair because not all staff are treated alike. In countries where evaluators have some leeway, the conditions under which inspections are carried out can differ considerably. More information should be provided on the standards and criteria used in the course of school evaluation and a minimum level of standardization assured.

151. *Evaluating the evaluators*. One possible way to reassure those being assessed is to ensure that evaluators are themselves evaluated. Most countries have internal evaluations of the inspection work. Internal evaluation of the inspectorate includes discussions on approaches and instruments within the inspectorate, often under the supervision of a co-ordinating inspector or a chief inspector. Data on the experiences of school managers or parents with the inspectorate can be systematically gathered, as in the United Kingdom and the Netherlands. In the Czech Republic, after finalising the inspection report, the Czech School Inspectorate (ČŠI) Headquarters can send a questionnaire to the school heads in order to receive feedback on the work of the inspection in the school. A separate unit within the inspectorate can be exclusively focused on the quality of the inspection work (Standaert, 2001). In Scotland, an audit unit is responsible for evaluating the work of the inspectorate, including the results of the follow-up to the inspections, while a working group of inspectors is permanently engaged on the effectiveness of the guidelines (Standaert, 2001). Systematic evaluation of each inspector takes place in several countries. In the Flemish Community of Belgium, each member of the inspectorate receives an annual evaluation during the first three years of his career. Thereafter, an evaluation takes place at least every two years. The Department of Evaluation in the Swedish agency was evaluated in 2004, and the agency as a whole in 2005. Similarly, the Danish Evaluation Institute's methods were evaluated in 2005 by Högskoleverket, a Swedish institution usually evaluating the Swedish higher education sector.

152. Engaging in dialogue and consultations: If evaluation seeks to improve the quality of schools, it is important that their staff, which is in the best position to decide what should be evaluated, can play an active part in defining standards and designing the evaluation process. By contrast, when evaluation is used above all to report to the education authorities, or to supply information for the purposes of external evaluation, it is preferable for the authorities to partly determine the criteria. They will then have comparable data at their disposal (Eurydice, 2004). However, schools and school staff will accept more easily to be evaluated and to implement the recommendations formulated by the inspectorate if they are consulted in the design of the process. In addition to taking the opinions and claims of educational practitioners as regards the feasibility and relevance of evaluation schemes into account, the consultation of school staff recognises their skills and experience.

5. EMPIRICAL EVIDENCE ON THE EFFECTS OF SCHOOL EVALUATION

153. This section reviews empirical evidence on the impact of school evaluation schemes. It initially explains the difficulties in assessing the impact of school evaluation schemes. It then considers the mixed qualitative and quantitative empirical findings and analyses the circumstances under which school evaluation schemes seem to be more effective.

5.1 Difficulties in measuring the impact of school evaluation

154. *Data availability* The quality of research on the effects of school evaluation schemes is highly dependent on the quality of the available data. The existence of data, however, does not necessarily imply the ability of researchers to identify the impact of school evaluation *per se*. Within an educational system since several policy initiatives are typically introduced concurrently, it is often challenging to find appropriate methods to separate and identify the impact of a single policy initiative.

155. *Inference*: When reforms are implemented nationwide, it is by definition difficult to assess the impact of a specific evaluation scheme and to establish a counter-factual, since there is no control group. Döbert *et al*. (2004) report some of the typical obstacles to program evaluation in the case of England. A number of studies were carried out in England in the 1990s on the effects of local management of schools (LMS) on student outcomes but were unable to assess its effects: LMS was introduced nationally (so there was no control group of schools with no LMS to allow comparison); there was no national pupil assessment dataset on attainment by which effects over time could be tracked; several reforms were introduced at the same time (the national curriculum, assessment and OFSTED inspection) so that it would be intricate to disentangle the respective causal impact of different programs. Identifying whether specific evaluation schemes have a causal impact on pupil performance requires comparability of student data and sufficient variation of school evaluation across schools. The latter seldom exists as school evaluation policies are generally defined nationally, so there are no treatment and control schools within a given country (Wößmann *et al*., 2007). The endogeneity of school institutional arrangements in the absence of experiments with random assignment of evaluation schemes makes it difficult to measure the causal impact of evaluation schemes on pupil outcomes.

5.2 Empirical evidence on school evaluation schemes

5.2.1 Empirical evidence on the impact of external evaluation systems on school performance

Effects of accountability threats on low-performing schools

156. An emerging body of evidence provides indications that the threat of sanctions on low-performing schools can raise student test scores in the short run. However, the extent to which these test score improvements are due to schools' gaming of the accountability system is unclear. Two strands of research emerge as to the question of whether the observed test score gains from accountability pressure are primarily the result of gaming behaviour or reflect the impact of educational reforms: one such strand suggests positive, short run effects of sanction threats on test scores, and another strand documents the existence of school gaming behaviour under accountability pressure. A number of studies on Florida's

accountability system have found that sanction threats have raised the observed test scores of students during the time that they are attending the schools 'under threat' (Figlio and Rouse, 2006). The introduction of school accountability in Chicago has been shown to raise test scores by a greater extent within schools at higher risk of being placed on probation (Jacob, 2005). A second strand of research has found that schools under accountability pressure may alter testing conditions or student classifications to boost observed test scores in ways unintended by policymakers. For instance, pressured schools have been shown to remove low-achieving students from school rating calculations by reclassifying them into special education (Jacob, 2005). Schools under pressure have also been observed to manipulate testing conditions through teacher cheating (Jacob and Levitt, 2003).

Impact of external evaluation systems on improvement strategies and school management

157. Chiang (2009) provides evidence that suggests accountability systems have a positive impact on student outcomes. First, using a regression discontinuity design that exploits Florida's system of imposing sanction threats on the basis of a cutoff level of performance, he finds that sanction threats on elementary schools have a persistent, positive impact on student test scores even after the affected students have progressed into middle school. Threat induced math improvements from elementary school largely persist at least through the first 1 to 2 years of middle school, while evidence for persistence of improvements in reading is less consistent. Second, estimates of immediate expenditure responses to the onset of sanctions are found to induce significant changes in the quantity, allocation, or use of educational inputs within threatened schools: sanction threats are found to increase school spending on instructional technology, curricular development, and teacher training. However, the analysis cannot necessarily be generalized to higher performing schools (Chiang, 2009). Similarly, surveys of Florida's principals conducted by Rouse *et al*. (2006) show that sanction threats induce schools to lengthen instructional time, change school wide schedules, and increase the planning time and professional development with which teachers can improve instruction. These findings are consistent with the claim that schools facing accountability pressure are likely to implement pedagogical and curricular changes in their attempts to raise student achievement. By contrast, Jacob (2003) examines changes in school budgetary line items over a five-year period spanning the introduction of accountability in Chicago and finds declines in fine arts spending, decreases in the ratio of aides to teachers, and increases in the ratio of supervisors to teachers within lower-performing schools relative to higher performing schools.

Effects of public release of results and school "rankings" on school results

158. There is some evidence that schools develop interventions to improve the performance of some groups of students (in detriment of others) in response to the short-run incentives created by school ratings systems in the United States. For instance, Reback (2007) examines whether school accountability systems that use test score measures based only on minimum competency, such as those created under *No Child Left Behind* in the United States, influence the distribution of student achievement. Because school ratings in these systems only incorporate students' test scores via pass rates, this type of system increases incentives for schools to improve the performance of students who are on the margin of passing while offering no short-run incentives for schools to improve other students' performance. Schools might therefore concentrate on the marginal students, to the detriment of very low achieving students and high achieving students. Texas' accountability program has been using student performance measures to assign discrete grades or ratings to all schools and/or school districts for several years. Using student level panel data from Texas during the 1990's, Reback estimates schools' short-run incentives to improve the performance of different groups of students. The empirical results suggest that schools respond to the accountability system by taking actions which influence the distribution of student achievement. These actions appear to be both broad measures that help all low achieving students, as well as more targeted measures to assist the students whose performance is critical to the schools' accountability ratings. Additional distributional effects are apparent for the same school across different years. When a school has

a greater short-run incentive to raise a pass rate, the performance of very low achieving students increases even if these students have a negligible chance of passing. In contrast, relatively high achieving students perform worse than usual if their own performance is irrelevant to the short run accountability incentives. However, other studies have found that statewide accountability programs have not led to reductions in high school dropout rates or increases in the rate of college attendance (Carnoy *et al*., 2003). An explanation that reconciles these findings is that schools have been raising the achievement of students who are marginal in terms of passing the state exam, and these types of students remain likely to graduate high school on schedule but unlikely to go to college (Reback, 2007).

5.2.2 Empirical evidence on the impact of evaluation schemes for student performance

Impact of external school evaluation upon student performance

159. Although there is a vast literature on school effectiveness and school improvement, empirically based studies of the effects of school inspections on school performance remain rare. An empirical study by Rosenthal (2004) investigates the impact of Ofsted inspections on the exam performance of students of inspected schools. The paper finds no evidence that the occurrence of an Ofsted visit has beneficial effects on the exam results of the school's students. The results indicate a small *negative* direct effect on exam results. The author suggests that perhaps the efforts by teaching staff in responding to the demands of the school inspection are great enough to divert resources from teaching so as to affect adversely pupil achievement in the year of the visit. According to the author, the requirements of the inspections may even crowd out the school's own improvement plans for the school's activities over the critical period. Hence, beneficial efficiency gains resulting from the program must be weighed against the potential costs of inspections, including potentially disruptive effects on pupils' education.

Impact of performance-based funding accountability systems on student results

160. Some evidence from U.S. studies suggests that performance-based funding accountability programs may have a positive impact on student outcomes. For instance, Ladd (1999) focuses on measuring the effects of Dallas's accountability program on student outcomes, as measured by pass rates on the state criterion-referenced test. During the early 1990s, the Dallas Independent School District engaged in the use of a performance-based accountability system based on financial incentives to school staff to increase student learning. Using a panel data set for schools in large Texas' cities, Ladd's study measures the gains in student performance in Dallas relative to those in other cities. Positive and relatively large effects are found for Hispanic and white seventh graders, but not for black students. Other potentially positive changes include the fall in the Dallas drop-out rate relative to that for other cities and the fact that principals are being replaced more readily than in the past, which could signify real internal change in the system. However, little attention is paid to processes within schools. According to Ladd, the next step would be to determine first whether the gains in student performance on the test translate into gains in student learning, and second to relate the gains in outcomes to information about how principals and teachers changed their behaviour in response to the accountability system and how those changes affected outcomes. An additional component would require a closer look at the potentially unintended effects of such programs.

Effects of school leaders' involvement in evaluation on student outcomes

161. Witziers, Bosker and Krüger (2003) find a positive relationship between student outcomes and several dimensions of leadership, such as supervision and evaluation, monitoring, and defining and communicating mission. The authors use a quantitative meta-analysis to estimate the effect size of educational leadership on student achievement among multinational cross-sectional research reports conducted between 1986 and 1996. One should nevertheless take caution to interpret the results in terms of

causes and effects: given the fact that school effects are by definition small (achievement differences are best explained by student characteristics), samples of schools tend to be too small to detect significant effects.

Impact of the publication of schools' student achievement data upon student performance

162. Evidence from PISA research suggests that the public release of student results has a positive impact upon student performance. Evaluation and accountability information was collected in PISA 2006 and analyzed to measure the impact of accountability systems upon student performance. The publication of schools' student achievement data was found to have a statistically significant positive impact upon student performance even after accounting for demographic and socioeconomic background characteristics and other school institutional or program characteristics. Fifteen-year-old students in schools that published student achievement data scored, on average, 3.5 score points higher on the PISA science scale than students in schools that did not publish achievement data, all other things being equal (OECD, 2007c).

5.2.3 Effects of school evaluation on teachers and school leaders

Effects of school evaluation on teacher effectiveness and motivation

163. Accountability-oriented reforms may shift teachers' motivational goals: teachers' intrinsic motivation to teach may be partly displaced by such extrinsic goals as money and "winning" the school's ranking competition when the stakes are high. As described earlier, evidence from a qualitative study in the Canadian province of Ontario makes it clear that those teachers who are constrained in ways likely to reduce their own intrinsic motivation to teach may behave in more controlling ways and be less effective in teaching their students (Leithwood *et al*., 2000). The study aimed to understand the sources of motivation influencing teachers' and school administrators' implementation of a government performance-based approach to school reform relying on the use of a common provincial report card in all schools aligned to the curriculum frameworks and student standards. Student testing was initiated five years earlier and the results released publicly, giving the media the ability to compare schools and to publish school rankings within districts. The majority of respondents perceived the effects of accountability policies to be mostly negative. Evidence about educators' personal goals indicated that because the majority of respondents believed that the government's intentions for many of its policies were unrelated to improving teaching and learning, they found little that resonated with their own professional goals. However, a small majority of administrators did associate some of the government's policies with goals close to their own. Even though nonalignment with teachers' goals was the dominant finding, the motivation to implement government policies might have been significantly enhanced had stronger connections been made with such motives as improving the quality of teaching and learning rather than only reducing the budget for education and pursuing political ends. The context in which these judgments were formed was characterized by conflict and hostility between educators and the government, suggesting that government approaches to policy making have an important influence on the success of policy implementation. Lack of moral authority on the government's part, growing out of a perceived discrepancy in goals, is one important reason for this attitude toward the government. Evidence from this case study makes it clear that not only do teachers believe they are primarily accountable to students and parents but also that they are likely to rank the government last among the agencies to whom they feel they should be accountable.

Effects of school evaluation on school leadership

164. Findings from a small scale empirical study of teacher leadership in the United Kingdom provide evidence of the effect of external accountability mechanisms on school leadership (Muijs, Harris 2006). Qualitative evidence was collected from ten schools, purposively selected as being sites where teacher leadership was operational and considered to be contributing to improvement. Data was collected through

interviews with school staff, including teacher leaders, classroom teachers, managers and head teachers. External accountability mechanisms, especially in low performing schools, were found to be a barrier to the development of teacher leadership. Indeed, these mechanisms put a strong burden on teachers and on senior management that makes the distribution of leadership more difficult and riskier. The proliferation of top-down initiatives emanating from central government was viewed as stifling teacher initiative and leadership capabilities.

6. CONCLUDING REMARKS

165. As a response to the demands for high quality education, a growing emphasis is being placed upon comprehensive school evaluation systems as they are central to school improvement efforts, systems of school accountability and school choice. A consensus has emerged on the need for clarifying the purpose of evaluation emphasized and on the importance of including a broad set of stakeholders in the design and implementation of the evaluation process. Effective evaluation schemes require the acceptance and support of those being evaluated, whose positions and claims have to be taken into account throughout the development of the evaluation process.

166. As a general rule, countries make little use of the great variety of methods and approaches that are available for the evaluation of schools. Although the cost of implementing comprehensive school evaluations is not trivial, most researchers recommend broadening the range of aspects to be evaluated and the range of actors to be mobilized. Also, in addition to the traditional inspection-based external evaluation and internal evaluation procedures, a number of other arrangements could be adopted: peer reviews, self-assessments, or the involvement of a more diverse set of evaluators (e.g. experts, critical friends, parents) would contribute to evaluating schools more thoroughly. Moreover, better articulation between school evaluation and other components of the school system's evaluation framework and better alignment between evaluation procedures led by different agencies (municipalities, school providers, inspectorates) would help prevent inconsistency of objectives and strengthen the coherence of the evaluation system.

167. As regards the consequences of school evaluation for school improvement, connection between the formative and summative functions of evaluation is often deemed insufficient. Similarly, it seems clear that more technical support and training is needed to allow schools to use the findings from external evaluation in a way conducive to school improvement.

REFERENCES

The descriptive information contained in this report was drawn from the UNESCO International Bureau of Education's "world data on education", Eurydice data base, national ministries of education websites and the OECD's education data base (Country studies, Education at a Glance). The descriptive part of the literature review is based primarily on publicly accessible documents, such as evaluation policies and evaluation reports, including the websites of organizations performing evaluations. Academic papers are indicated below.

Alvik, T. (1996), "Self Evaluation: What, Why, How, by Whom, for Whom", Collaborative Project Self-evaluation in School Development, Dundee, CIDREE.

Assessment Reform Group (2002), *Testing, Motivation and Learning*, Faculty of Education, University of Cambridge.

Barrett, B. (2009), "No Child Left Behind and the assault on teachers' professional practices and identities*", Teaching and Teacher Education*, pp. 1–8.

Bénabou, R., F. Kramarz and C. Prost (2004), "*Zone d'éducation prioritaire: quels moyens pour quels résultats?*", *Économie et statistique*, No. 380.

Bhola, D., J. Impara and W. Buckendahl (2003), "Aligning tests with states' content standards: Methods and issues", *Educational Measurement: Issues and practice*, Vol. 22, No. 3, pp. 21-29.

Bishop, J. (2006), "Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn" Forthcoming in: Eric A. Hanushek, Finis Welch (eds.), *Handbook of the Economics of Education*, Amsterdam.

Bishop J., J. Moriarty and F. Mane (2000), "Diplomas for learning, not seat time: the impacts of New York Regents examinations", *Economics of Education Review,* 19, pp. 333–349.

Björklund A., P.-A. Edin, P. Freriksson and A. Krueger (2004), "Education, Equality and Efficiency: An Analysis of Swedish School Reforms during the 1990s", *IFAU Report* 2004:1, Uppsala: Institute for Labour Market Policy Evaluation.

Bowen G., R. Rose and W. Ware (2006), "The Reliability and Validity of the School Success Profile Learning Organization Measure", *Evaluation and Program Plannin.g,* No. 29, pp. 97–104.

Campbell C. and L. Ben (2008), "Using data to support educational improvement", *Educational assessment, evaluation and accountability*, No. 21, pp. 47-65.

Carnoy, M., S. Loeb and T. Smith (2003), "The impact of accountability policies in Texas high schools", in M. Carnoy, R. Elmore and L. Siskin (eds.), *The New Accountability: High Schools and High-Stakes Testing*, Routledge Falmer.

Chiang, H. (2009), "How accountability pressure on failing schools affects student achievement" *Journal of Public Economics,* No. 93, pp. 1045–1057.

Chrispeels, J. and K. Martin (2002), "Four school leadership teams define their roles within organisational and political structures to improve student learning", *School Effectiveness and School improvement,* Vol. 13, No. 3, pp. 327-365.

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

Döbert H., E. Klieme and W. Sroka (eds), (2004), *Conditions of school performance in seven countries: a quest for understanding the international variation of PISA results*, Waxmann Verlag GmbH.

Eraut, M. (1978), "Accountability at school level-some options of their implications", in T. Becher and S. MacLure (eds.), Accountability in education, NFER: Nelon, Windor and Slough, pp. 152-199.

Eurydice (2004), "Evaluation of schools providing compulsory education", Brussels: Eurydice.

Eurydice (2007) "School autonomy in Europe: Policies and measures", European Commission, Brussels,

Eurydice (2008), "Levels of autonomy and responsibility of teachers in Europe", European Commission, Brussels, Education DG.

Figlio D. and C. Rouse (2006),"Do accountability and voucher threats improve low-performing schools?", *Journal of Public Economics* No. 90, pp. 239–255.

Freiberg, H.J. (Ed.) (1999), *School climate: Measuring, improving and sustaining healthy learning environments.* London: Falmer.

Goldstein, H. (2001), "Using pupil performance data for judging schools and teachers: Scope and limitations" *British Educational Research Journal, 2*7, pp. 433-442.

Griffiths J., L. Vidovich and A. Chapman (2008), "Outcomes approaches to assessment: comparing non-government and government case-study schools in Western Australia", *The Curriculum Journal*, Vol. 19, No. 3, pp. 161-175.

Gurr, D. (2007), "Diversity and progress in school accountability systems in Australia", *Educational Research on Policy Practice*, No. 6, pp. 165-186.

Hægeland T., L. Kirkebøen, O. Raaum and K. Salvanes (2004), "Marks across lower secondary schools in Norway What can be explained by the composition of pupils and school resources?", Report 2004/11, *Statistics Norway*.

Hallinger, P. and R. Heck (1998),"Exploring the principals' contribution to school effectiveness: 1980-1995". *School Effectiveness and School Improvement*, *9*(2), pp. 157-191.

Hamilton, L. and M. Berends (April 2006), "Instructional Practices Related to Standards and Assessments", Document No. WR-374-EDU.

Hamilton L., M. Berends and B. Stecher (2005), "Teachers' responses to Standards-Based Accountability", Rand Education.

Hamilton, L. and B. Stecher (2006), "Measuring Educators' Responses to High-Stakes testing", Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, pp. 7-11.

Hamilton, L., B.M. Stecher, S.P. Klein (2002), *Making sense of test-based accountability in education*, Rand education.

Hanushek, Eric A. (2006), Some U.S. Evidence on How the Distribution of Educational Outcomes Can Be Changed. Forthcoming in: Paul E. Peterson, Ludger Wößmann (eds.), *Schools and the Equal Opportunity Problem*. Cambridge, MA: MIT Press.

Hanushek E. and M. Raymond (2005), "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, Vol. 24, No. 2, pp. 297–327.

Jacob, B. (2003), "Getting inside accountability: lessons from Chicago", in W. Gale and J. Pack (eds.), *Brookings–Wharton Papers on Urban Affairs*, Brookings Institution Press, Washington, D.C., pp. 41–70.

Jacob, B. (2005), "Accountability, incentives, and behavior: the impact of high-stakes testing in the Chicago public schools", *Journal of Public Economics* No. 89, pp. 761–796.

Jacob, B. and S. Levitt (2003),"Rotten apples: an investigation of the prevalence and predictors of teacher cheating. Quarterly Journal of Economics 118, pp. 843–877.

Janssens F., G. van Amelsvoort (2008), "School self-evaluations and school inspections in Europe: An exploratory study", Studies *in Educational Evaluation,* No. 34, pp. 15–23.

Kochan, S (2007), "Evolution of school performance research in the USA: from school effectiveness to school accountability and back", *International Handbook of school effectiveness and improvement*, T. Townsend (eds.), Springer International Handbooks of Education.

Koretz D., D. Caffrey and L. Hamilton (2001), "Towards a framework for validating gains under high-stakes conditions", CRESST/Harvard Graduate school for education University of California.

Kyriakides, L. and R. Campbell (2004), "School self-evaluation and school improvement: A critique of values and procedures", *Studies in Educational Evaluation*, 30(1), pp. 23–37.

Ladd, H. (1999), "The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes", *Economics of Education Review,* No. 18, pp. 1–16.

Ladd, H. and R. Walsh (2002), "Implementing value-added measures of school effectiveness: getting the incentives right", *Economics of Education Review,* No. 2, pp. 1 1–17.

Learmonth, J. (2000), *Inspection: what's in it for schools*? London, Routledge Falmer.

Leithwood, K. (1992), "The move to transformational leadership", *Educational Leadership*, 49 (5), pp. 8-12.

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

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

Mac Beath, J., Meuret, D., Schratz, M. and L. Jakobsen (1999), *Evaluating quality in school education. A European pilot project*, Final report, Brussels: European Commission, Education, Training and Youth Directorate.

Martin C. and E. Guzman (2005), "Performance evaluation or standardized testing of aptitudes? Innovations at the margins Mexico's of school system" *International Journal of Educational Development,* No. 25, pp. 145–155.

McBeath, J. and K. Myers (2002), *Self-evaluation: What's in it for schools?* London: Routledge/ Falmer.

McDonald, A. (2001), "The Prevalence and Effects of Test Anxiety in School Children", *Educational Psychology*, Vol. 21, No. 1, pp. 89-101.

Monsen, L.I. (2002), "School-Based Evaluation in Norway: Why Is It So Difficult to Convince Teachers of its Usefulness?" in D. Nevo (ed.), *School-Based Evaluation: An International Perspective*, JAI Press, Oxford, pp. 73-88.

Muijs D. and A. Harris (2006), "Teacher led school improvement: Teacher leadership in the UK", *Teaching and Teacher Education* No. 22, pp. 961–972.

Nevo, D. (2001), "School evaluation: internal or external?", *Studies in Educational Evaluation*, 27(2), pp. 95–106.

Nevo D. (2002), "School-based evaluation: an international perspective*", Advances in program evaluation*, Vol. 8, Elsevier Science Ltd, Oxford, UK.

Northouse, P. (2001), *Leadership: Theory and practice,* Thousand Oaks, CA: Sage.

OECD (1996). *Evaluating and Reforming Education Systems*, OECD, Paris.

OECD (1999), *Making Education Count: Developing and Using International Indicators*, OECD, Paris.

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

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

OECD (2007a), *No more Failures: Ten Steps to Equity in Education*, Education and training policy, Simon Field, Malgorzata Kuczera, Beatriz Pont, OECD, Paris.

OECD (2007b), "School accountability, autonomy, choice and the level of student achievement : International evidence from PISA 2003", OECD Education Working Papers No. 13, L. Wöbmann, E. Lüdemann, G. Schütz and M. West.

OECD (2007c), *PISA 2006: Science Competencies for Tomorrow's World,* OECD, Paris.

OECD (2008a), *Education at a Glance: OECD Indicators*, OECD, Paris.

OECD (2008b), "How do OECD countries take stock of progress and performance in education systems? Evidence and issues", *Education Policy Committee Ad Hoc Workshop-Taking stock of education performance: from student testing to system evaluation,* OECD, Paris.

OECD (2008c), *Improving school leadership*, Vol. 1 & 2, B. Pont, D. Nusche and D. Hopkins(eds.), OECD, Paris.

OECD (2009), "Assessment and innovation in education", OECD Education Working Paper No. 24, Janet Looney, OECD, Paris.

OFSTED (2005), Annual Report of Her Majesty's Chief Inspector of Schools: standards and quality in education 2002/03, www.ofsted.gov.uk

OFSTED (2008), The Annual Report of Her Majesty's Chief Inspection, www.ofsted.gov.uk/publications/annualreport0708

Randall. R. (2008), "Teaching to the rating: School accountability and the distribution of student achievement", *Journal of Public Economics*, No. 92, pp. 1394–1415.

Rochkind J., A. Ott, J. Immerwahr, J. Doble, and J. Johnson (2007), *Lessons Learned: New Teachers Talk about Their Jobs, Challenges, and Long-Range Plans: A Report from the National Comprehensive Center for Teacher Quality and Public Agenda* , New York: Public Agenda.

Rosenthal, L. (2004), "Do school inspections improve school quality? Ofsted inspections and school examination results in the UK", *Economics of Education Review,* No. 23, pp. 143–151.

Rowe, K. (2005), "Evidence for the kinds of feedback data that support both student and teacher learning", *Australian Council for Educational Research,* Background paper to address presented at the Research Conference Lumina Grand Hyatt Hotel, Melbourne, 7-9 August 2005.

Rutter, M. and B. Maughan (2002), "School Effectiveness Findings 1979–2002", *Journal of School Psychology*, Vol. 40, No. 6, pp. 451 – 475.

Sackney, L. (2007), "A history of school effectiveness and improvement research in Canada over the past 25 years", International *Handbook of school effectiveness and improvement*, T. Townsend (eds.), *Springer International Handbooks of Education*.

Sammons, P., K. Elliot, W. Welcomme, B. Taggart and R. Levacic (2004), "England", in H. Döbert, E. Klieme and W. Sroka (eds.), *Conditions of school performance in seven countries: a quest for understanding the international variation of PISA results*, Waxmann Verlag GmbH.

Sanders, W. and S. Horn (1994), "The Tennessee value-added assessment system (TVASS): Mixed model methodology in educational assessment", *Journal of Personnel Evaluation in Education,* No. 8, pp. 299-311.

Scheerens, J. and R. Bosker (1997), *The fundations for educational effectiveness,* Oxford: Pergamon.

Scheerens J., G. van Amelsvoort and C. Donoughue (1999), "Aspects of the organizational and political context of school evaluation in four European countries", *Studies in Educational Evaluation,* No. 25, pp. 79-108.

Sims, D (2008), "Strategic responses to school accountability measures:It's all in the timing", *Economics of Education Review* , No. 27, pp. 58–68.

Skolverket (2005), "Educational inspection 2004, Summary of inspection results", A summary of report 266, The Swedish National agency for Education.

Standaert, R (2001), *Inspectorates of education in Europe: a critical analysis,* Acco, Leuven, Belgium.

Stecher, B. M. (2002), "Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practices", in L. Hamilton, B. Stecher and S. Klein (eds.), *Making Sense of Test-Based Accountability in Education*, RAND, Santa Monica, CA, pp. 79-100.

Stufflebeam, D. and A. Shinkfield (1990), *Systematic evaluation,* Lancaster: Kluwer-Nijhoff.

Teddlie, C. and S. Stringfield (2007), "A history of school effectiveness and improvement research in the USA focusing on the past quarter century", in T. Townsend (ed.), *International Handbook of school effectiveness and improvement*, Springer International Handbooks of Education.

Toch, T. and R. Rothman (2008), "Rush to Judgment: Teacher Evaluation in Public Education", *Education sector reports*.

Townsend, T. (Ed.) (2007), *International Handbook of school effectiveness and improvement*, Springer International Handbooks of Education.

Ungerleider, C. and B. Levin (2007), "Accountability, funding and school improvement in Canada", in T. Townsend (Ed.), *International Handbook of school effectiveness and improvement*, Springer International Handbooks of Education.

van Bruggen, J. (2001), "Functions of inspectorates in Europe", Paper presented at the International Inspection, Academy, Berlin.

Vanhoof J., P. van Petegem and S. De Maeyer (2009), "Attitudes towards school self-evaluation", *Studies in Educational Evaluation,* No. 35, pp. 21–28.

Visscher, A., S. Karsten, T. de Jong and R. Bosker (2000), "Evidence on the intended and unintended effects of publishing school performance indicators", *Evaluation and Research in Education,* No. 14, pp. 254-267.

West, M. and D. Hopkins (1997), *Using evaluation data to improve the quality of schooling*, Frankfurt, Germany: ECER-Conference.

Wilson, M., P. Hallman, R. Pecheone and P. Moss (2007), "Using Student Achievement Test Scores as Evidence of External Validity for Indicators of Teacher Quality: Connecticut's Beginning Educator Support and Training Program" (unpublished paper).

Witziers, B., R. Bosker and M. Krüger (2003),"Educational Leadership and Student Achievement: The Elusive Search for an Association", *Educational Administration Quarterly*, No. 39.

Woessmann, L. (2006), "Efficiency and equity of European education and training policies", *CESifo*, Working paper No. 1779.

Zimmerman, B. and M. Dibenedetto (2008), "Mastery learning and assessment: implications for students and teachers in an era of high-stakes testing", *Psychology in the schools*, Vol. 45(3).

EXISTING OECD EDUCATION WORKING PAPERS

No. 1 *Teacher Demand and Supply: Improving Teaching Quality and Addressing Teacher Shortages* (2002), Paulo Santiago.

No. 2 *Teacher Education and the Teaching Career in an Era of Lifelong Learning* (2002), John Coolahan.

No. 3 *Towards an Understanding of the Mechanisms That Link Qualifications and Lifelong Learning* (2003), Friederike Behringer and Mike Coles.

No. 4 *Measuring Educational Productivity in Standards-Based Accountability Systems: Introducing the SES Return on Spending Index* (2005), Martin Hampel.

No. 5 *PISA 2000: Sample Weight Problems in Austria* (2006), Erich Neuwirth.

No. 6 *Funding Systems and their Effects on Higher Education Systems – International Report* (2007), Franz Strehl, Sabine Reisinger and Michael Kalatschan.

No. 7 *On the Edge: Securing a Sustainable Future for Higher Education* (2007), OECD/IMHE-HEFCE.

No. 8 *Globalisation and Higher Education* (2007), Simon Margison and Marijk van der Wende.

No. 9 *Understanding the Regional Contriubution of Higher Education Institutions: A Literature Review* (2007), Peter Arbo and Paul Benneworth.

No. 10 *Effects of Tertiary Expansion – Crowding-out Effects and Labour Market Matches for the Higher Educated* (2007), Bo Hansson.

No. 11 *Skilled Voices? Reflections on Political Participation and Education in Austria* (2007), Florian Walter and Sieglinde K. Rosenberger.

No. 12 *Education and Civic Engagement: Review of Research and a Study on Norwegian Youths* (2007), Jon Lauglo and Tormod Oia.

No. 13 *School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003* (2007), Ludger Wössmann, Elke Lüdemann, Gabriela Schütz and Martin R. West.

No. 14 *School Accountability, Autonomy, Choice, and the Equity of Student Achievement: International Evidence from PISA 2003* (2007), Gabriela Schütz, Martin R. West, Ludger Wössmann.

No. 15 *Assessment of learning outcomes in higher education: a comparative review of selected practices* (2008), Deborah Nusche.

No. 16 *Approaches and Challenges to Capital Funding for Educational Facilities* (2008), Ann Gorey.

No. 17 *Recent Developments in Intellectual Capital Reporting and their Policy Implications* (2008), W. Richard Frederick.

No. 18 *Employers' Perspectives on the Roles of Human Capital Development and Management in Creating Value* (2008), L. Bassi and D. McMurrer.

No. 19 *Job-related Training and Benefits for Individuals: A Review of evidence and explanations* (2008), Bo Hansson.

No. 20 *A Framework for Monitoring Transition Systems* (2008), Rolf van der Velden.

No. 21 *Final Report of the Development of an International Adult Learning Module (OECD AL Module)* (2008), Bo Hansson and Helmut Kuwan.

No. 22 *What Works in Migrant Education? A Review of Evidence and Policy Options* (2009), Deborah Nusche.

No. 23 *Teacher Evaluation: Current Practices in OECD Countries and a Literature Review* (2009), Marlène Isoré.

No. 24 *Assessment and Innovation in Education* (2009), Janet Looney.

No. 25 *Do quasi-markets foster innovation in education? A comparative perspective* (2009), Christopher Lubienksi.

No. 26 *International Adult Literacy and Basic Skills Surveys in the OECD Area* (2009), William Thorn.

No. 27 *Learning Organisations* (2009), Nathalie Greenan and Edward Lorenz.

No. 28 *Les compétences bien sûr, mais pas seulement. L'influence des compétences cognitives en lecture sur la participation aux études postsecondaires chez les jeunes Canadiens* (2009).

No. 29 *Impact of proficiency on early entrants to the labour market: evidence from the YITS* (2009).

No. 30 *How does academic ability affect educational and labour market pathways in Canada?* (2009).

No. 31 *Does reading proficiency at age 15 affect pathways through learning and work?* (2009).

No. 32 *PISA and TIMSS: similarities and differences* (2009).

No. 33 *PIAAC Reading Components: A Conceptual Framework* (2009), John P. Sabatini and Kelly M. Bruce.

No. 34 *PIAAC Literacy: A Conceptual Framework* (2009), PIAAC Literacy Expert Group.

No. 35 *PIAAC Numeracy: A Conceptual Framework* (2009), PIAAC Numeracy Expert Group.

No. 36 *PIAAC Problem Solving in Technology-Rich Environments: A Conceptual Framework* (2009), PIAAC Expert Group on Problem Solving in Technology-Rich Environments.

No. 37 *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: What Skills will be Measured by the Programme for the International Assessment of Adult Competencies (PIAAC)?* (2009), William Thorn.

No. 38 *ICT in Initial Teacher Training: Research Review* (2009), Ann-Britt Enochsson and Caroline Rizza.

No. 39 *Education and Obsesity in Four OECD Countries* (2009), Franco Sassi, Marion Devaux, Jody Church, Michele Cecchini and Francesca Borgonovi.

No. 40 *Estimating the Marginal Effects of Education: on Health and Civic and Social Engagement: A Feasibility Study* (2009), Don Kenkel.

No. 41 *21st century skills and competences for New Millennium Learners in OECD countries* (2009), Katerina Ananiadou and Magdalena Claro.

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, SourceOECD: www.sourceoecd.org

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.