Chapter 3

The evaluation and assessment framework:

Embracing a holistic approach

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

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

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

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

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

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

Definitions

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

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

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

Scope

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

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

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

The educational context shapes the evaluation and assessment framework

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

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

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

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

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

Governance

Objectives and functions of evaluation and assessment

Improving student outcomes is the ultimate objective of evaluation and assessment

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

**Evaluation and assessment have a range of functions**

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

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

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

In Mexico it is apparent that the policy initiatives in evaluation and assessment of the last few years have emphasised accountability over development. For instance, the in-service teacher appraisal system currently in place is mostly focussed on salary progression and rewards and places less emphasis on its links to professional development, the assessment of students is oriented towards summative scores, and school evaluation is
essentially reduced to accountability through the publication of student standardised assessment results (Santiago et al., 2012b). By contrast, in New Zealand, the development function of evaluation and assessment is strongly emphasised. In primary schools, student assessment is mostly formative and provides detailed feedback rather than assigning numerical marks. The sample-based standardised assessments do not carry high stakes for students. Assessment in secondary schools is more summative but there are opportunities for schools to reassess and resubmit internal assessments to maximise learner success and students also receive their marked external assessments back. A range of tools and professional development offers are available for teachers to help them gather a variety of evidence of student learning to allow nuanced overall judgements on performance. Also, New Zealand’s approach to school evaluation has evolved to focus attention on building the capacity of schools for effective self-evaluation and strategic planning for improvement of teaching and learning (Nusche et al., 2012).

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

<table>
<thead>
<tr>
<th>Use of results for accountability</th>
<th>Use of results for development</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Australia, Chile</td>
<td>Mexico, Slovak Republic, Sweden</td>
</tr>
<tr>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Belgium (Fi.), Canada, Israel, Korea, New Zealand</td>
<td>Czech Republic, France, Hungary, Ireland, Netherlands, Poland, Portugal</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Denmark, Iceland, Norway</td>
<td>Austria, Belgium (Fr.), Estonia, Finland, Italy, Luxembourg, Slovenia, Spain</td>
</tr>
</tbody>
</table>

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

Source: Taxonomy developed using information supplied by countries on features of evaluation and assessment frameworks (synthesised in Chapters 4 to 8) and information from Education at a Glance (OECD, 2011).
A key challenge therefore is to find the right balance between the accountability and the development functions of evaluation and assessment. While transparency of information, high-quality data, and the accountability of school agents are essential for a well-functioning evaluation and assessment system, it is important to ensure that the existing data and information are actually used for development and improvement. This requires reflection on designing mechanisms to ensure that the results of evaluation and assessment activities feed back into teaching and learning practices, school improvement, and education policy development.

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

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

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

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

In other cases, the development function of evaluation and assessment is less well understood by school agents. In Norway, according to an evaluation of the National Quality Assessment System (NKVS), among stakeholders there is no clear understanding of the whole system for evaluation and assessment (Allerup et al., 2009). The evaluation showed that the key elements of NKVS were understood to be the national tests, user surveys, inspections and international tests. This reflects that the more accountability-oriented elements of the evaluation and assessment framework are more prevalent in the perceptions of school agents than the support and guidance tools developed by the Directorate for Education and Training for local use and analysis. This is in spite of the fact that the proposal for the creation of the Directorate had clearly stated that “quality
assessment should primarily be a tool to be used by teachers, schools and students in their quality development work” (Nusche et al., 2011b).

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

**Responsibilities within the evaluation and assessment framework**

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

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

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

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

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

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

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

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

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

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

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

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

The creation of some of these agencies was triggered by the establishment of national education standards and the need for their monitoring. For example, in Austria, the establishment of the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) was associated with the development, implementation and monitoring of education standards. Since its creation, BIFIE assumes responsibility for the preparation of the annual “National Education Report” (Specht and Sobanski, 2012). Similarly, in the Slovak Republic, the National Institute of Certified Measurement (NÚCEM) takes responsibility for measuring and evaluating the quality of education. In Italy, the National Institute for the Evaluation of the Education System (INVALSI) has been given the functional responsibility of system evaluation under the supervision of the National Ministry for Education, University and Research.
### Table 3.2 Specialised intermediate agencies with a role in the evaluation and assessment framework, by country (2012)

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

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

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

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

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

In highly decentralised countries, there are typically excellent quality assurance initiatives at the local and school level but a number of challenges arise. The school owners (e.g. regions, municipalities and private providers) generally vary in their capacity for and commitment to evaluation and assessment activities. The instruments used for quality assurance and reference standards might be extremely diverse. They are also often not documented at the national level. There might also be few mechanisms to identify good practice and share it across the entire system. In Hungary, school maintainers (local governments and independent maintainers) take responsibility for evaluating the effectiveness of the pedagogical work of schools and their professionals. However, no
criteria are defined at the national level and there is great autonomy to decide the specific approach followed. Public schools are required to follow the rules stipulated in the Local Council Quality Management Programme (Hungarian Ministry of Education and Culture, 2010). In the Flemish Community of Belgium, the regulatory framework for quality assurance is developed to fit the Freedom of education principle, which grants the schools the right to develop their own concept of quality of education. This entails investments in the capacity of schools concerning school self-evaluations, use of data by school agents, and the development of school development plans (Shewbridge et al, 2011a).

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

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

Integration of the non-public sector

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

The integration of non-public schools in the overall evaluation and assessment framework varies considerably across countries. In Belgium, where the Freedom of education principle provides parents with free school choice, the evaluation and assessment framework treats all schools equally, i.e. all schools are subject to the same rules and
engage similarly in centrally dictated evaluation and assessment activities (in addition to the their own sector’s specific initiatives). By contrast, in Chile, which places strong emphasis on parental freedom of school choice through its voucher programme, central requirements for evaluation and assessment in private schools are considerably more limited than in public schools. For instance, the approach for teacher appraisal in the private school sector (most of which is publicly subsidised) consists of giving independence to school providers to run their own procedures. However, private schools are required to participate in national standardised student assessments (Santiago et al., forthcoming).

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

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

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

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

General goals for the education system

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

Overall goals for education systems typically emphasise the following aspects:

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

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

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

For instance, in Mexico, the General Education Law provides clear goals for education emphasising the development of individuals and the promotion of values and attitudes. These are associated with broader social and economic goals. Statements about the aims for the education system such as its promotion of diversity, equity and quality and its role in developing successful learners and informed citizens are articulated. In addition, federal governments in office establish priorities for educational policy, which provide the framework for policy development. For instance, one of the six objectives of the 2007-2012 Education Sector Programme was “to promote the development and use of information and communication technologies in the education system in order to provide support for student learning, increase student life abilities and favour student entry into the knowledge society”. Education targets to be achieved by 2012 were also established with associated indicators to permit the monitoring of their achievement (e.g. in primary
school, a proportion of 82% and 83% students achieving at the basic proficiency level in national tests, in Spanish and mathematics respectively) (SEP and INEE, 2011).

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

**Equity goals in education systems**

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

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

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

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

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

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

The development of student learning objectives is a complex exercise. For example, as explained in Looney (2011a), standards writers may have difficulty agreeing on the knowledge and skills that are most important. While the majority of OECD countries now promote skills for “learning-to-learn”, including skills for problem-solving, critical analysis, as well as supporting students in developing greater autonomy, and so on, there may still be deep-seated tensions about the goals of education. Such “culture wars” (Finn and Kanstoroom, 2001), may lead to the development of standards that are vague (thereby avoiding controversy), or at the other extreme, standards that are overly detailed, making
it difficult to identify priorities for learning, and providing little useful guidance for instruction or the development of assessments (Chudowsky and Pellegrino, 2003).

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

**Alignment between goals for student learning and evaluation and assessment**

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

In a well-aligned system, student learning objectives are comprehensive, and clearly define the content and cognitive processes students are expected to demonstrate at different stages. If systems are poorly aligned, it is impossible for the system, schools and teachers to draw valid conclusions about student performance, or to adapt teaching to better meet identified needs (Linn, 2001, 2005; Haertel and Herman, 2005). Misalignment can have serious consequences on instruction and learning (Alliance for Excellent Education, 2010). For instance, one negative consequence of disconnected education standards and student standardised assessment is “teaching to the test”, where teachers may emphasise assessment taking skills and low-level content, rather than important learning goals expressed by the standards (Alliance for Excellent Education, 2010). Also, if standards are poorly designed or not specific enough, teachers are more likely to focus on tests, thus narrowing the focus of teaching (Stecher et al., 2000).
Box 3.1 The Comprehensive Reform of Basic Education in Mexico

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

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

Source: Santiago et al. (2012b).

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

Alignment issues also apply to external school evaluation, school self-evaluation, teacher appraisal and school leader appraisal. References for school evaluation, school leader appraisal and teacher appraisal need to align with student learning objectives and include the associated criteria to be used by evaluators. Agencies in charge of school evaluation, school leader appraisal and teacher appraisal should also be able to provide evidence of inter-evaluator rating reliability. In the context of school self-evaluations, staff may need to achieve consensus regarding goals for the evaluation, and the criteria by which they will judge school performance. All these evaluation activities require good “social alignment”, i.e. social capital in systems, including shared values, motives and efforts.
around educational goals and the principles underlining them (Baker, 2004; Hargreaves, 2003). In socially aligned systems, institutions and actors work together to define challenges and to consider alternative courses of action. This alignment is vital for system learning and improvement (Looney, 2011a).

**Design**

**Principles**

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

**Placing students at the centre of the evaluation and assessment framework**

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

By contrast, in some other countries, teaching, learning and assessment still take place in a somewhat “traditional” setting with the teacher leading his/her classroom, the students typically not involved in the planning and organisation of lessons and assessment concentrating on summative scores. In such settings, while teachers may also focus on the development of skills for critical analysis and reasoning, students have few opportunities to experiment with or apply new knowledge in different situations. Knowledge is presented within traditional, discrete categories (language, mathematics, science, arts, and so on), and there may be few opportunities to link learning across subjects. There is little
attention paid to student motivation, or to developing skills for “learning to learn” (OECD, 2009). In these countries, the opportunity given to parents and students to influence student learning also tends to be more limited and there is little emphasis on the development of students’ own capacity to regulate their learning. Other practices which might require further strengthening in these countries are the communication of learning expectations to students, the opportunities for performance feedback and mechanisms for individualised support. Also, collecting the views and perspectives of parents and students to inform school improvement through the systematic use of surveys is not a general practice in some countries. This includes surveys designed by teachers to collect student views on their teaching.

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

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

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

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

**Focussing on student outcomes**

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

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

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

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

Committing to transparency through the reporting of evaluation and assessment results

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

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

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

In Australia, following the Melbourne Declaration on Educational Goals for Young Australians in December 2008 Ministers agreed that public reporting on Australian schools would: support improving performance and school outcomes; be both locally and nationally relevant; and be timely, consistent and comparable. In June 2009, they agreed a set of eight principles and related protocols for reporting on schooling in Australia, the Principles and Protocols for Reporting on Schooling in Australia (MCEECDYA, 2009). This is a document which makes clear their commitment to transparent accountability. The principles relate directly to data on student outcomes and information about the school context and resourcing. The protocols are designed to promote the integrity of the
process and to provide safeguards against simplistic comparisons being made amongst schools (Santiago et al., 2011).

**Figure 3.3 Publication of school evaluations (2007-08)**

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

![Bar chart showing the publication of school evaluations](http://dx.doi.org/10.1787/888932791172)


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

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

In some countries, classroom observations are not a systematic part of the appraisal of each teacher or are not conducted in the context of school evaluation and there is little tradition for school leaders or teacher peers to conduct informal classroom observations with an evaluative focus. Figure 3.5 displays perceptions by school agents on the use of
the direct appraisal of classroom teaching as a criterion of teacher appraisal and school evaluation, as part of the 2007-08 TALIS study in lower secondary schools. Countries in which there is less stress on the direct appraisal of classroom teaching include Australia, Denmark, Iceland, Norway, Portugal and Spain.

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

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

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


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

The ability for the evaluation and assessment framework to effect changes in the classroom and improve student learning depends to a great extent on its reliance on teacher professionalism. In New Zealand, the evaluation and assessment system is grounded in a strong belief in teacher professionalism. Teachers are seen as the main experts not only in teaching but also in assessing their students. This is in contrast to some other countries where student assessment is conceived to a great extent as an activity separate from teaching and undertaken by school-external psychometric experts. In New Zealand, the assessment of national standards is not based on whole cohort standardised assessment but instead relies on teachers’ capacity to assess against the standards. National standards aim to provide external reference points of expected student performance while leaving the responsibility for choosing assessment methods and forming overall judgements with teachers. The approach to national monitoring (based on sample-based assessments) also involves teachers in the assessment activities. A range of teacher professional development programmes, as well as mentoring and induction for new teachers, aim to ensure strong teacher competencies in assessment. Teacher professionalism is further supported by well-established approaches to teacher appraisal. Teachers have a good degree of ownership of the appraisal process. It is the professional...
body of teachers (New Zealand Teachers Council), and not an external agency that has taken the lead role in defining standards for teacher registration. Individual teachers are actively involved in their appraisal processes (both for registration and for performance management) through self-assessment of their own practices (Nusche et al., 2012).

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

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

**Responding to diverse learner needs**

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

In Ireland, the identification of students with special educational needs is a key aspect of the assessment framework for primary and secondary education (Irish Department of Education and Skills, 2012). In New Zealand, much work has been undertaken to develop assessment tools and approaches that are adapted to different learner groups. For example, the Ministry of Education is working with Māori assessment experts to develop approaches for monitoring student outcomes in the context of the curriculum and the standards used in Māori-medium settings. There is also a focus on developing guidance and resources for teachers to develop narrative assessment approaches and Individual Education Programmes for students with special educational needs. A number of language and literacy assessment tools are also available to provide adequate assessment opportunities for English language learners. There has also been some focus on including attention to Māori learner needs in teacher standards and teacher appraisal procedures.
The Registered Teacher Criteria emphasise the bicultural context of New Zealand. School review practices have also been adapted to ensure that school reviews fulfil the commitment of the education sector to improving education outcomes for Māori and Pasifika students. System evaluation focuses attention on ensuring that information is collected not only on the whole group of students but also on specific groups, and in particular the Māori and Pasifika students. This is intended to provide relevant information to identify strategies to respond to diverse learning needs (Nusche et al., 2012).

Main components

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

- **Student assessment.** Student assessment in general includes national (full-cohort or sample-based) standardised assessments, with diagnostic and monitoring purposes, and externally based summative assessment, in particular in view of assessing students for secondary education certification. At the school level, student assessment plays the key role in informing schools and teachers about students’ individual achievement through teacher-based summative and formative assessments.

- **Teacher appraisal.** Procedures vary considerably across countries but, in addition to probationary processes, typically occur in two specific instances: (1) as part of performance management processes, including regular appraisal, to gain and maintain registration/accreditation to teach, and for promotion; and (2) to identify a select number of high-performing teachers to reward and acknowledge their teaching competence and performance. These formal schemes are often complemented with more informal school-level practices of feedback to teachers.

- **School evaluation.** There are generally two main forms of evaluation: school self-evaluation and school external evaluation or review. The latter is typically conducted by an external agency and involves a sequence of activities which may begin with self-reflection by the school, includes a visit by an external evaluator or team of evaluators, and leads to a summative report which may be published and may require a follow-up process. Schools may also be held accountable on the basis of comparable measures of student results.

- **The appraisal of school leaders.** Procedures vary considerably across countries but, in addition to probationary processes, are typically part of the employer’s performance management processes with emphasis on administrative and pedagogical leadership.

- **System evaluation, the evaluation of sub-national education systems, and the evaluation of a programme or a policy.** In general, education system evaluation involves: (i) the monitoring of student outcomes at a given point in time, including differences among different regions within the education system and given student groups (e.g. by gender, socio-economic or immigrant background); (ii) the monitoring of changes in student outcomes over time; (iii) the monitoring of the impact of given policy initiatives or educational programmes; (iv) the monitoring of demographic, administrative and contextual data which are useful to explain the outcomes of the education system; (v) the development of means through which the relevant information is provided to the different agents in the education system; and (vi) the use of the generated information for analysis, development and implementation of policies.
A challenge in some countries is that some key components of a comprehensive evaluation and assessment framework are underdeveloped or do not exist. For example, the formative assessment of students by teachers might be underdeveloped as a result of the focus placed on results and a classroom practice dominated by examination and test preparation. Teacher appraisal might also not be systematic, school self-evaluation might be undertaken with little capacity, and the appraisal of school leaders might be at an incipient stage of development. Figure 3.7 provides an indication of the extent of teacher appraisal and school evaluation procedures in countries which participated in the 2007-08 round of TALIS according to the perceptions of teachers and school principals. It suggests that in some countries such as Ireland, Italy, Portugal and Spain, teacher appraisal and school evaluation were not systematic practices across schools at the time the TALIS survey was conducted. The situation may have changed in the case of these countries since then. For example, post-primary schools in Ireland are now regularly evaluated using a range of inspection models and Portugal is in the process of implementing a teacher appraisal model.

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

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

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

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

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

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

Main elements within evaluation and assessment components

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

SYNERGIES FOR BETTER LEARNING: AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT © OECD 2013
• **Governance: Who?** This aspect concerns the overall design and organisation of one particular component of the evaluation and assessment framework. This includes the setting of the respective requirements and the distribution of responsibilities for the design of such evaluation component. It also includes the strategic goals of the particular evaluation component, its contribution to the overall evaluation and assessment framework and the balance between developmental and accountability functions.

• **Capabilities to assess and to use feedback: By Whom?** This aspect concerns the preparation to evaluate, to be evaluated and to use the results of an evaluation as well as the choice of the groups undertaking these functions. It includes issues such as: the choice of the evaluators and the development of the skills to perform an assessment; the preparation to be the subject of an evaluation; the development of competencies to effectively use the results of an evaluation for the improvement of practice; and the design of agencies to review evaluation results with a view to hold agents accountable and to inform policy development.

• **Scope: What?** Some evaluation processes concentrate on analyses of outcomes such as the level and distribution of students’ results. Other evaluation initiatives centre on processes such as the quality of teaching and learning or the effectiveness of school leadership.

• **Evaluation “technology”: How?** This aspect refers to the organisation of particular types of evaluation, that is the mix of instruments, criteria and standards, purposes, skills, and scope which are used to undertake a given evaluation or assessment. For instance, a teacher appraisal model may be based on a range of instruments such as self-appraisal, classroom observation and a teacher portfolio; be focussed on the teaching and learning process; be undertaken in relation to reference standards for the teaching profession; have both development and accountability purposes; and be based on experienced peers. Hence, this aspect refers to the way different aspects are combined to produce a given evaluation model.

• **Purposes: For what?** This encompasses the objectives of a particular evaluation process and the mechanisms designed to ensure that evaluation results are used in a way such objectives are reached. The objectives of an evaluation process typically consist of development and accountability. Examples of mechanisms to use evaluation results include performance feedback, professional development plans, financial and other rewards, publication of results to the general public, and policy adjustments.

• **Agents involved: With whom?** This mostly deals with the political economy of reform aspects of evaluation and assessment procedures. It relates to the involvement of a range of stakeholders such as parents, students, teachers, school leaders, teacher unions, educational administrators and policy makers in the development and implementation of evaluation and assessment processes.

This conceptual framework is used in subsequent chapters to analyse each of the components of the evaluation and assessment framework.
3. THE EVALUATION AND ASSESSMENT FRAMEWORK: EMBRACING A HOLISTIC APPROACH

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

**By whom?**
- Capabilities to assess and to use feedback

**Who?**
- Unit assessed – Governance

**How?**
- Evaluation “technology”
  - Methodology and procedures: mix of instruments, criteria, purposes, and skills to assess a given unit
  - Mapping of different kinds of feedback to different units
  - Interrelation between different types of evaluation

**What?**
- Scope/elements assessed

**With whom?**
- Agents involved
  - Students, teachers, and school leaders
  - Educational administrators and policy makers
  - Parents, communities, taxpayers
  - Teacher unions, educators, education professionals

**For what?**
- Purposes
  - Accountability
  - Development

**Mechanisms to use feedback**
- Performance feedback
- Professional development/formative implications
- Financial and other implications/recognition and reward
- Sanctions
- Reporting
- Information/publication of results
- Policy adjustments/development

**Criteria and standards**
- Outcome objectives
- Reference standards
- Performance criteria

**Instruments**
- Student national examination
- Student national assessment
- Teacher-based student assessment
- Teaching classroom observation
- Teacher self-appraisal
- Teacher portfolio
- School self-evaluation
- School external evaluation
- International student assessment
- Performance indicators
- Surveys

**Inputs**
- Infrastructure
- Number and characteristics of staff and students

**Processes**
- Teaching and learning
- School leadership
- Educational administration

**Outcomes**
- Level and distribution of students’ results
- Equity of student results

**Users of feedback**
- Whom? Teachers, school leaders, educational administrators, policy makers
- Skills and know-how

**Assessed units**
- Whom? Teachers, school leaders, educational administrators, policy makers
- Skills and know-how

**Evaluators**
- Whom? Inspection, peers, school leaders, agencies
- Skills and know-how

**Evaluation agencies**
- Agencies to review results of evaluation and assessments
- Skills and know-how

**Evaluation and assessment**
- An international perspective on evaluation and assessment

SYNERGIES FOR BETTER LEARNING: AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT © OECD 2013
Articulations

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

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

Figure 3.9 Articulations within the evaluation and assessment framework
Articulations within components

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

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

<table>
<thead>
<tr>
<th>Component</th>
<th>Articulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student assessment</strong></td>
<td>- Alignment between student learning objectives and student assessment</td>
</tr>
<tr>
<td></td>
<td>- Linkages between student summative assessment and student formative assessment</td>
</tr>
<tr>
<td></td>
<td>- Linkages between classroom-based assessment and external assessment</td>
</tr>
<tr>
<td></td>
<td>- Alignment between performance ratings (marks) and educational standards</td>
</tr>
<tr>
<td></td>
<td>- Alignment between student assessment and assessment courses in teacher education</td>
</tr>
<tr>
<td><strong>Teacher appraisal</strong></td>
<td>- Alignment between teaching standards and student learning objectives</td>
</tr>
<tr>
<td></td>
<td>- Alignment between teaching standards and teacher appraisal</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between teacher appraisal and professional development</td>
</tr>
<tr>
<td></td>
<td>- Alignment between teaching standards, registration processes and career structure</td>
</tr>
<tr>
<td></td>
<td>- Articulation between school-based teacher appraisal and externally driven teacher appraisal</td>
</tr>
<tr>
<td></td>
<td>- Linkages between formative teacher appraisal and high-stakes teacher appraisal</td>
</tr>
<tr>
<td></td>
<td>- Alignment between skills taught in teacher education and teaching standards assessed in teacher appraisal</td>
</tr>
<tr>
<td><strong>School evaluation</strong></td>
<td>- Alignment between nationally agreed criteria for school quality and student learning objectives</td>
</tr>
<tr>
<td></td>
<td>- Alignment between nationally agreed criteria for school quality and school evaluation</td>
</tr>
<tr>
<td></td>
<td>- Articulation between school self-evaluation and external school evaluation</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between school evaluation and school development</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between school evaluation and school reporting</td>
</tr>
<tr>
<td></td>
<td>- Complementarity between school performance measures and school review or inspection</td>
</tr>
<tr>
<td></td>
<td>- Articulation between distinct school evaluation processes conducted by education authorities at different levels</td>
</tr>
<tr>
<td><strong>School leader appraisal</strong></td>
<td>- Alignment between school leadership standards and student learning objectives</td>
</tr>
<tr>
<td></td>
<td>- Alignment between school leadership standards and school leader appraisal</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between school leader appraisal and professional development</td>
</tr>
<tr>
<td></td>
<td>- Alignment between school leadership standards, registration processes and career structure</td>
</tr>
<tr>
<td></td>
<td>- Linkages between formative school leader appraisal and high-stakes school leader appraisal</td>
</tr>
<tr>
<td><strong>Education system evaluation</strong></td>
<td>- Alignment between education goals, student learning objectives and education system evaluation</td>
</tr>
<tr>
<td></td>
<td>- Placing education system evaluation in the broader context of public services evaluation</td>
</tr>
<tr>
<td></td>
<td>- Mapping between education goals and targets and measures of education system evaluation</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between education system evaluation and education policy development</td>
</tr>
<tr>
<td></td>
<td>- Systematic linkages between education system evaluation, public reporting and information systems</td>
</tr>
<tr>
<td></td>
<td>- Coordination between the evaluation of the education system and the evaluation of education sub-systems</td>
</tr>
<tr>
<td></td>
<td>- Complementarity between quantitative measures and qualitative system evaluation</td>
</tr>
<tr>
<td></td>
<td>- Articulation between full-cohort and sample-based standardised assessments for national monitoring</td>
</tr>
</tbody>
</table>
An example of articulation within an evaluation component is the complementarity between school self-evaluation and school external evaluation. If the two processes are developed separately, this may lead to a costly duplication of data gathering and evaluation processes and the potential for external evaluation to test, affirm, strengthen and broaden school self-evaluation is not realised. In addition, the role that external evaluation can play in helping schools build their evaluative capacity and report progress effectively is also limited.

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

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

Articulations between components

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

An example of articulation across evaluation components is the mutually reinforcing linkage between school evaluation and teacher appraisal. This relates to a range of aspects such as: school-based teacher appraisal being validated by school evaluation processes; making the focus of school evaluation on teacher effectiveness systematic across schools; and school development processes exploring links to the evaluation of teaching practice. Another example is the articulation between school evaluation and the appraisal of school leaders. The results of school evaluation can usefully inform the appraisal of school leaders given that it reviews the performance of the school led by the leader being appraised and typically includes the assessment of leadership.
### Table 3.5 Key articulations between components of the evaluation and assessment framework

<table>
<thead>
<tr>
<th>Component</th>
<th>Articulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student assessment</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Teacher appraisal | Student assessment results to inform teacher appraisal  
Teacher’s skills for student assessment to be systematically reviewed in teacher appraisal  
Teacher’s skills to use student assessment data to be reviewed in teacher appraisal  
Teacher appraisal results to inform professional development on competencies for student assessment |
| School evaluation | School evaluation to review student assessment practices in schools, including school-based student assessment criteria and teachers’ skills for student assessment  
School evaluation to review school’s ability to moderate marking within the school and involve teachers in collaborative work around student assessment  
School evaluation to review school’s capacity to analyse and use student assessment data  
Student assessment results to inform school evaluation  
School reporting to include student assessment results  
School evaluation results to inform school capacity development on competencies for student assessment |
| School leader appraisal | Student assessment results to inform school leader appraisal  
School leader’s skills for implementing student assessment practices in the school to be systematically reviewed in school leader appraisal  
School leader appraisal results to inform professional development on skills for evaluation and assessment |
| Education system evaluation | Student assessment as part of the national monitoring of student outcomes, namely through standardised national student assessment  
Educational trajectories of students as part of system evaluation (longitudinal studies)  
Secondary analyses of student assessment data, including differentiated analyses across student subgroups  
Results of education system evaluation to inform both the approaches to the national monitoring of student outcomes and the competencies to be assessed  
Standardised national student assessment to assist the moderation of classroom-based student assessment  
Policy evaluation to assess coherence of classroom-based student assessment across schools and its alignment with student learning objectives |
| **Teacher appraisal** | |
| School evaluation | Validation of school-based teacher appraisal by external school evaluation  
Review of school’s application of external teacher appraisal by school evaluation processes  
Focus of school evaluation on teacher effectiveness to be systematic across schools  
School development processes, including school self-evaluation to use results of teacher appraisal  
Results of school evaluation to inform the development of school-based teacher appraisal  
Results of teacher appraisal to inform external school evaluation |
| School leader appraisal | Teacher appraisal and school leader appraisal frameworks to inform each other  
School leader appraisal to assess role of school leader in teacher appraisal, including the capacity to provide individual professional feedback and to lead the school’s strategies to improve teacher effectiveness  
Results of teacher appraisal to inform school leader appraisal |
| Education system evaluation | Evaluation at the system and sub-system levels to use the information generated by teacher appraisal  
Policy evaluation to assess consistency of teacher appraisal across schools  
Results of education system evaluation to inform both the approaches to teacher appraisal and the competencies to be assessed by teacher appraisal |
Moderation processes

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

Linkages to the classroom

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

In some countries, the focus has been on structures, procedures, programmes and resources defined at the central level but, while these components of policy are clearly important, there has been a less clear articulation of ways for the national agenda for education to generate improvements in classroom practice through the assessment and
evaluation procedures which are closer to the place of learning. A strong top-down national vision for evaluation and assessment might constrain the ownership of evaluation and assessment procedures by school agents and result in a greater focus on the accountability function. This might also reflect the greater technical capacity at the centre and the more limited competencies at the local level to engage in evaluation and assessment activities. However, establishing links between evaluation and assessment and classroom learning inevitably requires establishing clear roles for local structures – school management, school supervision, local education authorities – in the implementation of evaluation and assessment policies. The point is that the fulfilment of the developmental function of evaluation and assessment requires articulation at the local level.

**Capacity for evaluation and assessment**

*Competencies for evaluation and assessment*

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

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

- teacher capacity to assess against the whole range of curriculum goals to ensure consistency of marking across schools
- teacher capacity for formative assessment
- data handling skills of school agents (e.g. use of results from student standardised assessments)
- information for parents and other stakeholders to gain a good understanding of some outcome reporting
- capacity for taking on the role of external evaluator (e.g. in school evaluation, teacher appraisal)
- evaluation competencies of groups or agencies undertaking evaluation activities such as school evaluation or teacher appraisal, including school governing boards
- standardised assessment development, educational measurement, psychometrics, validation of test items, scaling methods
- externally based assessment (e.g. national examinations for certification)
- analytical capacity for educational planning and policy development.
As an example, Figure 3.10 provides teachers’ perceptions of professional development needs in student assessment practices based on TALIS data for 2007-08. In OECD countries such as Italy, Korea, Norway and Slovenia, over 20% of teachers indicate that they have a high level of need for professional development in student assessment practices.

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

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

Another area of priority is capacity for formative assessment. For instance, in Ireland, whole-school evaluation reports by inspectors and the 2009 National Assessments survey (Eivers et al., 2010) indicate that there is still significant scope for development in terms of the formative use of student assessment not only to improve the learning of individual students in the classroom but to promote improvement at whole-school level. These concerns have been instrumental in informing the new emphasis placed on evaluation and assessment in Ireland’s 2011 National Literacy and Numeracy Strategy (Irish Department of Education and Skills, 2012).
A major concern in countries in which education governance is highly decentralised is the capacity of local education authorities to ensure the quality assurance of their schools. In many of these countries, it is recognised that many (often smaller) municipalities lack the capacity to develop and manage robust quality assurance systems, monitor schools effectively and follow up with schools accordingly. Further, the background and qualifications of officials in local education authorities responsible for school evaluation tend to vary significantly. This often leads to low expectations and the lack of an evaluation culture; variability and inconsistency in quality assurance practices; and limited capacity and skills of schools to use data.

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

**Evaluation advisors in Denmark**

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

Source: Shewbridge et al. (2011b).

**The National List of Experts in Hungary**

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


**Skills for school leadership**

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

In most countries, there is no specific initial education to train school leaders, nor does the specific career of school leader exist (see Chapter 7). Also, it is still common across the OECD area that school principals focus their work on administrative tasks, limiting their pedagogical leadership of schools. Figure 3.11 shows countries’ relative positions in terms of the perceptions of school principals of their management styles, using specific indices based on TALIS data. Countries where school principals perceive administrative tasks are predominant in their work include Ireland, Italy, Mexico, Norway and Turkey. By contrast, in countries such as Australia, Denmark, Hungary, Iceland, Korea, Poland and Slovenia instructional leadership seems to prevail over administrative tasks in the work of school principals, according to their perceptions. A large-scale
longitudinal study by Seashore Louis and colleagues (2010) based in the United States found that collective leadership focused on instructional improvement had a significant impact on teachers’ working relationships, and on student achievement.

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

Scores on TALIS administrative leadership and instructional leadership indices

![Graph showing scores on TALIS administrative leadership and instructional leadership indices for various countries.](http://dx.doi.org/10.1787/888932791286)

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


**Tools and guidelines for evaluation and assessment**

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

- detailed guidelines to implement the curriculum
- marking rubrics listing criteria for assessing and rating different aspects of student performance and exemplars illustrating student performance at different levels of achievement
- range of optional assessment tools for teachers to use in student assessment, including formative assessment
• instruments to interpret results in student standardised assessments
• tools for school self-evaluation
• guidance for the application of teacher appraisal, including instruments for self-appraisal.

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

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

Use of results

Knowledge management

Information systems

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

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

• standard framework for data collection and reporting
• data information/management systems – collection of data on students, teachers, schools and their performance over time
• public information – arrangements to share information about evaluation and assessment results with the general public

• identification of best practices and their dissemination across the system.

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

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

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

In Norway, the development of the School Portal (Skoleporten) has been instrumental in ensuring access for school owners and schools to monitoring information and analyses of their results. The School Portal is a web-based information tool presenting key education monitoring information including learning outcomes, learning environments, resources and basic school data. The Portal has an open part accessible to the general public and a password-protected part where schools and school owners can access more detailed information and benchmark themselves against the national average. This approach holds promise for encouraging a more systematic and well-integrated way of using analyses of data in the process of self-evaluation and improvement planning. At the same time, it provides the general public with information on educational outcomes
Similarly, in the French Community of Belgium, the Learning Portal (Le Portail de l’Enseignement) provides extensive information to the general public on student learning objectives, education indicators and education legislation as well as pedagogical tools and the instruments necessary for schools to engage in external evaluations (Blondin and Giot, 2011).

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

New South Wales SMART system

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

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


Source: DEEWR (2010).

Identification of best practices

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

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

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

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

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

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

Source: Nusche et al. (2011b).

**Innovation in education**

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

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

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

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

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

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

- **Innovative assessment as leading to pedagogical innovations**

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

**Evidence-based policy**

The principle of informing policies and educational practices with evidence from research, including that generated by using the results of evaluation and assessment activities, is among the main goals of the evaluation and assessment framework. The
The objective is to ensure that evaluation and assessment results are used in academic and policy research which subsequently informs the development of education policy. This includes developing evaluation and assessment policies which are evidence-based. Evidence-based policy includes aspects such as:

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

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

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

In Denmark, the Council for Evaluation and Quality Development of Primary and Lower Secondary Education, as part of its mandate to monitor the academic level in the Folkeskole (compulsory education), commissions research and evaluation studies. The latter often include studies on the implementation and use of new national evaluation and assessment tools, e.g. Individual Student Plans and municipal quality reports and have led
to considerations and pilots of how to make such tools most relevant to local needs. One of the ten major challenges identified by the Council for the “Folkeskole 2020” is to strengthen the systematic exchange of knowledge between research institutions and schools to promote school use of knowledge to improve teaching. The Council aims to collect and disseminate research results to support the formation of policies for school improvement (Danish Ministry of Education and Rambøll, 2011). In the French Community of Belgium, the “Monitoring Commission” (Commission de Pilotage), among other missions, reviews research in education and ensures the relevant results inform policy development and school practices. In addition, it identifies research gaps in education and proposes to the government a multi-year research plan establishing research priorities in education as well as objectives to be reached by research studies (Blondin and Giot, 2011).

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

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

Implementation of evaluation and assessment policies

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

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

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

Building consensus through consultation

A number of authors stress the importance of mechanisms to build consensus for successful adoption and implementation of policy initiatives (Fiske, 1996; Finlay et al., 1998; Corrales, 1999). There is extensive evidence that consensus is almost a prerequisite for successful implementation of policy reforms. As noted by Fiske (1996) with respect to school decentralisation, researchers are almost unanimous in arguing that if school decentralisation is going to be successfully carried out and have a positive impact on the quality of teaching and learning, it must be built on a foundation of broad consensus among the various actors involved and the various interest groups affected by such a change. And in fact, he observes on a basis of a comparative analysis that countries where leaders sought to build consensus for reform happen to be those where decentralisation was most successful. Building consensus is characterised by iterative processes of proposals and feedback which allow legitimate concerns to be taken into account, and hence reduce the likelihood of strong opposition by some stakeholder groups. A merit of structured consultations with stakeholders is that their regular involvement in policy design helps them build capacity over time. Another advantage is the potential for collective learning which might contribute to the development of a common concept of evaluation among stakeholders.
There is broad agreement in the literature that the involvement of stakeholders in education policy development cultivates a sense of joint ownership over policies, and hence helps build consensus over both the need and the relevance of reforms (Finlay et al., 1998; OECD, 2007). Policies promoting consensus build trust between the various stakeholder groups and policy makers. Keating (2011) analysed how various school districts in the United States developed and implemented new school principal appraisal systems. In most school districts, collaboration between different stakeholders (e.g. unions, teachers, school leaders and community representatives) played a key role in the design and implementation stages. The setting of shared priorities, negotiation, consensus building and transparency often resulted in greater ownership and acceptance among stakeholders.

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

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

In Finland, the objectives and priorities for educational evaluation are determined in the Education Evaluation Plan, which is devised by the Ministry of Education and Culture in collaboration with the Education Evaluation Council, the Higher Education Evaluation Council, the National Board of Education and other key groups. The members of the Education Evaluation Council represent the educational administration, education providers, teachers, students, employers, employees and researchers and thus can influence the aims and priorities of educational evaluation (Finnish Ministry of Education and Culture, forthcoming). In the French Community of Belgium a “Monitoring Commission” (Commission de Pilotage) has been given a key role in the monitoring of the education system. It has two main missions: it co-ordinates and reviews the coherence of the education system, and it follows the implementation of pedagogical reforms. Its membership reflects all the relevant actors in the education system: the school inspection, the school organisers, researchers, teacher unions and parent representatives (Blondin and Giot, 2011).
In Ireland, the involvement of stakeholders in the formulation of assessment and evaluation policy has been a prominent feature. A range of well-established frameworks promote dialogue and common action among the main stakeholders. The various committees that advise the assessment proposals of the National Council for Curriculum and Assessment (NCCA) and the national assessments of the Educational Research Centre (ERC) comprise members of relevant organisations and bodies including the Department of Education and Skills, school management groups, teacher unions, and parents’ groups (Irish Department of Education and Skills, 2012). Nevertheless, it should be noted that criticisms have been made of the way in which consultation and consensus building have slowed down the development of policy and the implementation of radical change (Irish Department of Education and Skills, 2012).

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

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

- A reference group was set up to guide the project “Strengthening of the evaluation culture in the *Folkeskole*”. The reference group, whose membership includes all the relevant stakeholder groups, meets on a regular basis to discuss the project. This includes, for instance, the development of the national student tests.

- The interest groups of the *Folkeskole* were involved in 2010-11 in a committee established by the Minister of education aiming at deregulating the *Folkeskole*.

- In 2007-08, the Danish Union of Teachers and the Ministry of Education collaborated on a project called “The School of the Month”. Each month, a school was celebrated for remarkable results. The project has since been pursued under the heading “the good example of the month” (www.skolestyrelsen.dk).

- The Local Government Denmark project “Partnership on the *Folkeskole*”, involving 34 municipalities, has been a platform for co-operation and reflection between municipalities. (www.kl.dk/ImageVault/Images/id_40353/ImageVaultHandler.aspx).

- The Quality and Supervision Agency in collaboration with the Danish Evaluation Institute carry out “inspirational seminars” for teachers and school pedagogical staff with a view of encouraging schools to develop evaluation activities.

- The Quality and Supervision Agency has all major stakeholder groups represented in focus groups, which are being summoned on a regular basis to provide input on different initiatives related to the strengthening of the evaluation culture in the *Folkeskole*.

- The different interest groups of the *Folkeskole* launch on a regular basis common actions and/or common proposals related to issues in the *Folkeskole*, e.g. a paper with the title Common knowledge – Common action.

With respect to the initiation of new policies, the combination of top-down and bottom-up initiatives is generally believed to foster consensus (Finlay et al., 1998). For instance, a study of evidence-informed policy making underlines how the involvement of practitioners – teachers, other educational staff and their unions – in the production of research evidence and in its interpretation and translation into policy gives them a strong sense of ownership and strengthens their confidence in the reform process (OECD, 2007).

In the Flemish Community of Belgium, the Flemish Education Council (VLOR) is an independent advisory body that the Ministry of Education and Training is required to consult when a draft decree is prepared for the Parliament. It brings together representatives from school organisers, school leaders, teachers, researchers, students and parents. One of the Council’s activities is the organisation of consultations and conferences to discuss the state of education, including the policy implications of analysis of the results of national student assessments (Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group, 2010).

**Involvement of professionals**

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

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

The involvement of teachers and school leaders in their own appraisal process has been identified as another key aspect for the successful implementation of individual appraisal processes. Engaging teachers and school leaders in their own appraisal, e.g. through setting objectives, self-appraisal and the preparation of individual portfolios, can help create a more effective and empowering process for teachers and school leaders, and, therefore, aid successful implementation.
Kennedy (2005) argues that highly dedicated teachers’ reform rejections do not come from their unwillingness to change or improve, but from “the sad fact that most reforms don’t acknowledge the realities of classroom teaching”. Also, imposed change is believed to create a “culture of compliance” (Datnow and Castellano, 2000; Leithwood et al., 2002). Teachers who are constrained in ways likely to reduce their own intrinsic motivation to teach may behave in more controlling ways and be less effective in teaching their students. By contrast, if teachers are involved in planning and implementing evaluation schemes, they are more likely to sustain reform efforts (Leithwood et al., 2000).

**Clarity of purposes**

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

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

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

**Evidence to inform consensus-building**

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

This highlights the importance of promoting research and making the evidence underlying the policy proposals available to the relevant stakeholders in helping convince practitioners and society at large. The objective is to raise awareness on problematic issues, enhance the national debate and disseminate evidence on the effectiveness and impact of different policy alternatives, and hence to find a consensus on educational evaluation policy. In the case of teacher appraisal, Milanowski and Heneman (2001) found that teachers’ overall favourableness toward a system newly implemented in a medium-sized school district in the United States was correlated with acceptance of the teaching standards, the perceived fairness of the process, the qualities of the evaluator, and the perception that the evaluation system has a positive impact on their teaching.
The concern of the Portuguese authorities to build teacher appraisal on research evidence and recognised good practice was a clear strength of the system when it was introduced. In 2007, the then Ministry of Education set up the Scientific Council for Teacher Evaluation (CCAP) as a consultative body to supervise and monitor the implementation of teacher appraisal (in late 2011, following the rationalisation of education services, the CCAP ceased its functions). The CCAP brought together educational researchers and distinguished teachers and as such was in a good position to recognise good evaluation practices, be informed of relevant research developments and provide evidence-based advice (Santiago et al., 2009).

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

Policy experimentation and pilots

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

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

Evaluation of implementation

Another approach is to periodically review and evaluate processes after the full implementation. Education professionals such as teachers and school leaders are more likely to accept a policy initiative today if they know that they will be able to express their concerns and provide advice on the necessary adjustments as the initiative evolves. Amsterdam et al. (2003) analysed the three-year development and validation of a school principal appraisal system (i.e. standards, criteria and instruments) in South Carolina, United States, that involved researchers from the South Carolina Educational Policy Center at the University of South Carolina, the South Carolina Department of Education,
a stakeholder committee (e.g. superintendents, school principals, teachers, guidance counsellors and journalists) and an expert panel. Superintendents responsible for carrying out the appraisal and school principals had the opportunity to further inform the development of standards and criteria through a survey and an online discussion group. The new standards, appraisal criteria and instruments subsequently underwent a process of piloting and validation through focus groups and a school principal survey of participants in the pilot to identify strengths and weaknesses of the new system. Based on their experiences, Amsterdam et al. argued that stakeholder input may help ensure that appraisal systems are practical and useful for those concerned, and that the appraisal is supported by key stakeholders. At the same time, the involvement of school leaders in the design of standards and appraisal criteria may help to establish an understanding of the aspects and criteria that school leaders will be appraised against.

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

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

**Capacity building**

Fiske (1996) underlines the importance of training policies for effective and successful implementation, as a means to ensure that all stakeholders are equipped and
prepared to take on the new roles and responsibilities that are required from them as a result of education reforms. Limited professional expertise of those with responsibility to evaluate and little analytical capacity to use the results from evaluation and assessment have the potential to harm the implementation of evaluation and assessment policies. For instance, in teacher appraisal models, it is fundamental to provide in-depth training to evaluators to guarantee that they are legitimate in the eyes of teachers. Also, scepticism towards data among educators resulting from a lack of capacity of schools and teachers to understand and use data effectively to inform development is also likely to increase implementation difficulties (Campbell and Levin, 2008).

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

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

**Resources**

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

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

**Timing**

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

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

**Pointers for future policy development**

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

Integrate the evaluation and assessment framework

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

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

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

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

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

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

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

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

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

Secure links to the classroom and draw on teacher professionalism

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

An important step in this direction could be a national reflection about the nature and purpose of evaluation components such as school evaluation, school leader appraisal, teacher appraisal and student formative assessment within the overall education reform strategy and the best approaches for these evaluation components to improve classroom practices. This reflection would shed light on strategies which can contribute to reinforce
the linkages between evaluation and assessment and classroom practice. Impacting classroom practice is likely to require the evaluation and assessment framework to place considerable emphasis on its developmental function. Channels which are likely to reinforce links to classroom practice include: an emphasis on teacher appraisal for the continuous improvement of teaching practices; ensuring teaching standards are aligned with student learning objectives; involving teachers in school evaluation, in particular through conceiving school self-evaluation as a collective process with responsibilities for teachers; ensuring that teachers are seen as the main experts not only in instructing but also in assessing their students, so teachers feel the ownership of student assessment and accept it as an integral part of teaching and learning; building teacher capacity for student formative assessment; and building teachers’ ability to assess against educational standards.

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

Give a prominent role to independent evaluation agencies

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

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

Promote national consistency while giving room for local diversity

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

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

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

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

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

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

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

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

Establish articulations between components of the evaluation and assessment framework

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

A prominent example is the articulation between school evaluation and teacher appraisal. Given that the systems of school evaluation and teacher appraisal and feedback have both the objective of maintaining standards and improving student performance, there are likely to be great benefits from the synergies between school evaluation and teacher appraisal. To achieve the greatest impact, the focus of school evaluation should either be linked to or have an effect on the focus of teacher appraisal. This indicates that school evaluation should comprise the monitoring of the quality of teaching and learning.
possibly include the external validation of school-based processes for teacher appraisal (holding school leaders accountable as necessary), and school development processes should explore links to the evaluation of teaching practice. In the context of school self-evaluation, it is also important to ensure the centrality of the evaluation of teaching quality and the feedback to individual teachers.

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

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

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

Build on some key principles to effectively implement evaluation and assessment

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

- **The centrality of teaching and learning**: It is critical to ensure that the evaluation of teaching and learning quality is central to the evaluation framework. Classroom observation should be a key element of teacher appraisal as well as an important instrument in external school evaluation. Similarly, the observation of teaching and feedback to individual teachers should be part of school self-evaluation processes. The effectiveness of the evaluation and assessment framework will depend to a great extent on the ability to cultivate a culture of sharing classroom practice, professional feedback and peer learning.

- **The importance of school leadership**: The effective operation of evaluation and assessment will depend to a great extent on the way the concept and practice of school leadership gains ground within the education system. It is difficult to envisage either effective teacher appraisal or productive school self review without strong leadership capacity. It is essential that school principals take direct
responsibility for exerting instructional leadership and for assuming the quality of education in their schools. Hence, the recruitment, appraisal, development and support for school leaders is of key importance in creating and sustaining effective evaluation and assessment cultures within schools.

- **Equity as a key dimension in the evaluation and assessment framework**: It is essential that evaluation and assessment contribute to advancing the equity goals of education systems. At the system level, it is imperative to identify educational disadvantage and understand its impact on student performance. Developing equity measures should be a priority in all countries. It is also important to ensure that evaluation and assessment procedures are fair to given groups such as cultural minorities and students with special needs.

- **A structure to integrate accountability and development**: The overall evaluation and assessment framework should include elements to accomplish both the accountability and developmental functions at all levels of the system (e.g. formative vs. summative assessment for students; professional development for teachers vs. career advancement decisions following teacher appraisal; data reporting vs. improvement action plans for schools) and provide a structure which can potentially integrate these two functions.

- **Commitment to transparency**: The overall evaluation and assessment framework can be strengthened by a high level of transparency in monitoring and publishing results.

**Capacity**

*Sustain efforts to improve capacity for evaluation and assessment*

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

There is also a need to reinforce the instructional leadership skills of school principals as their role in many countries still retains a more traditional focus on administrative tasks. The objective is that school leaders operate effective feedback, coaching and appraisal arrangements for their staff and effectively lead whole-school evaluation processes. This can primarily be achieved by redefining school leadership as educational leadership, and ensuring that the whole cohort of school leaders receives adequate training in “leadership for learning”. School leaders should be trained to implement an
authentic evaluation of teaching and learning, feedback and objective setting at their schools, including techniques in teacher observation (see also Chapters 5 and 7).

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

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

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

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

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

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

Maintain sound knowledge management within the overall evaluation and assessment framework

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

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

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

Commit to the use of evidence for policy development

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

Implementation

Anticipate potential implementation difficulties

The implementation of evaluation and assessment procedures requires the recognition of a range of important aspects. First, reaching agreements on the design of evaluation and assessment activities requires time for discussions and consultations with all stakeholders. Second, developing expertise in the system, including training evaluators is expensive and
requires time. Third, conducting evaluation processes induces additional workload for school agents. Fourth, aligning broader school reforms such as professional development opportunities with evaluation and assessment strategies requires more educational resources. It needs to be borne in mind that evaluation and the resulting feedback, reflection and development processes will only work if school agents make it work. To a great extent it is the motivated school agent who ensures the successful implementation of reforms in schools. Hence, it is imperative not only to find ways for school agents to identify with the goals and values of evaluation and assessment practices but also to ensure that such goals and values take account of school leader and teacher agency (OECD, 2006).

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

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

Communicate the rationale for reform

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

Use pilots before full implementation and review implementation

Policy experimentation and the use of pilots may also prove effective strategies to overcome blockages dictated by disagreements among stakeholders and to assess the
effectiveness of policy innovations before generalising them. Policy makers also need to ensure mechanisms and platforms for the ongoing review and development of evaluation and appraisal systems to ensure they are up-to-date with latest research and developments (e.g. through advisory or steering groups).

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

Ensure adequate capacity and sufficient resources

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

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

Notes

1. As of 2013, Hungary experienced a trend towards a larger degree of central decision-making in education. See endnote 1 in Chapter 2.

2. TALIS is the OECD’s Teaching and Learning International Survey, which was implemented in 2007-08, covering lower secondary education and with the participation of 23 countries (OECD, 2009). The results derived from TALIS are based on self-reports from teachers and principals and therefore represent their opinions, perceptions, beliefs and their accounts of their activities. Further information is available at www.oecd.org/edu/talis. The second cycle of TALIS (TALIS 2013) is being conducted in 2012-13.

3. Due to new regulations in Hungary on the role of education government in professional corporative bodies the Council for the Evaluation of Public Education ceased to exist legally in 2012.
References


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


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


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


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


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