Chapter 4

Student assessment:

Putting the learner at the centre

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

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

Introduction

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

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

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

Analytical approach

Definitions

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

Summative and formative assessment

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

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

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

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

Internal and external assessment

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

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

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

Key concepts related to student assessment

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

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

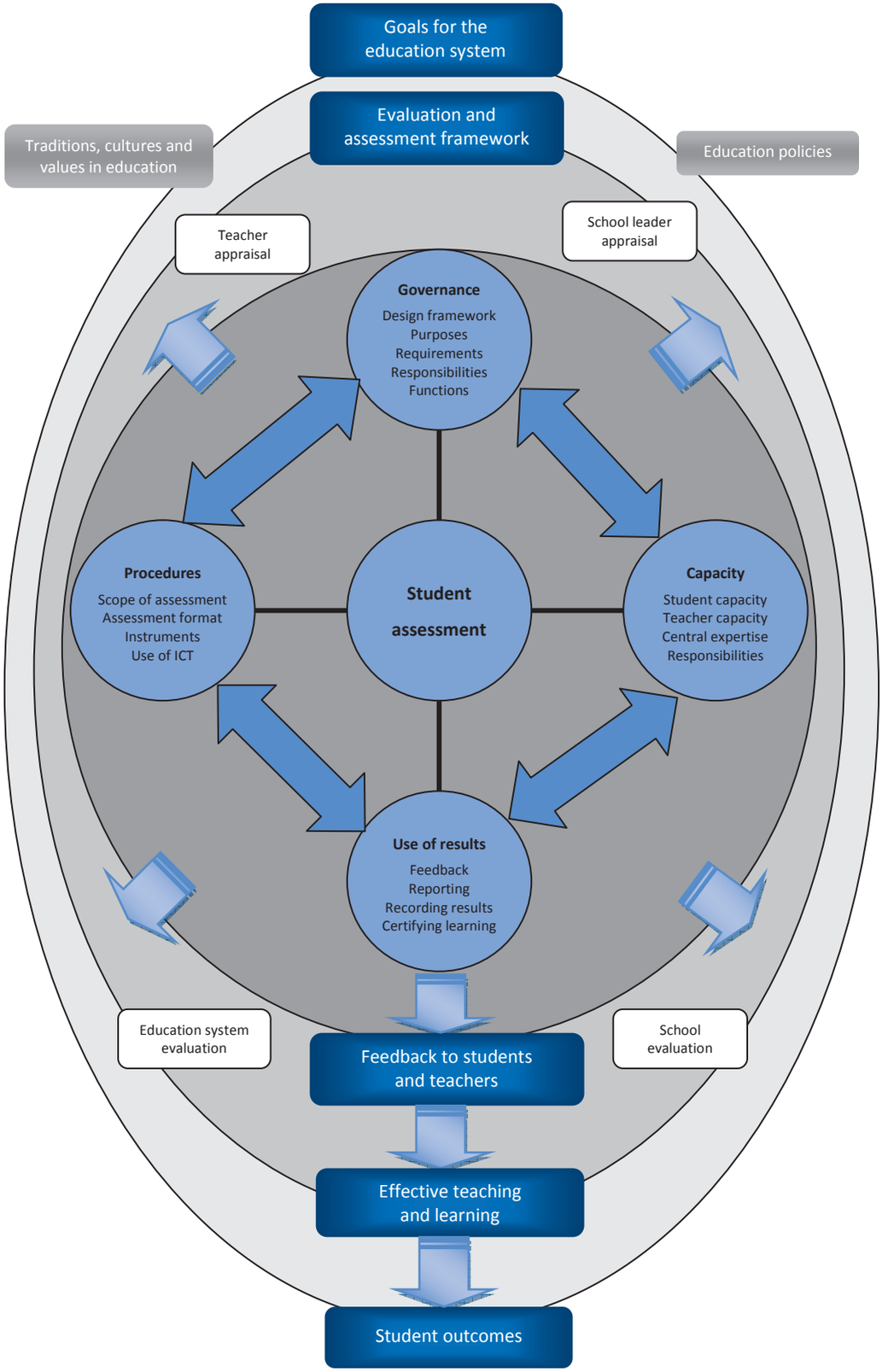
- **Transparency** relates to the degree to which information is available regarding expected learning outcomes, the criteria that will be used in judging student learning and the rules being applied when judgements are made about learning.
- **Usability** refers to how policy makers, school leaders,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

Conceptual framework

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

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

Figure 4.1 Conceptual framework for student assessment



Impact, drivers and contextual developments

The importance and impact of student assessment

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

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

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

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

Drivers and contextual developments

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

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

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

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

Formative classroom assessment

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

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

Summative classroom assessment

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

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

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

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

External assessments and examinations

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

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

Towards a new understanding of learning

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

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

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

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

Box 4.2 Key competencies around the world

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

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

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

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

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

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

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

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

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

Issues to be addressed

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

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

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

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

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

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

Towards a new understanding of assessment

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

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

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

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

Issues to be addressed

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

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

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

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

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

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

Governance

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

Purposes

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

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

Assessment for summative purposes

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

- **To motivate students** to increase effort and achievement. The marks, transcripts or diplomas connected to summative assessment are often conceived as rewards for having performed successfully on an assessment.
- **To provide information** about student performance to a range of different stakeholders, such as the students themselves, their parents, others within the school, or school-external players such as employers.
- **To select or group students** according to their achievement levels. In many countries, assessment results are used to stream students according to their ability levels within schools, or to select them into certain types of schools.
- **To certify learning and award qualifications** that grant students access to higher education institutions or certain professions.

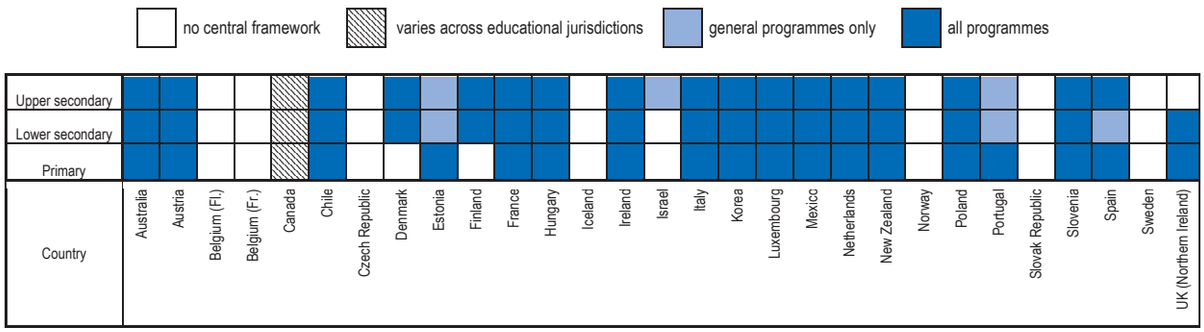
Across the OECD, countries draw on a range of different approaches to implement summative assessment. In the first years of education, summative assessment tends to be the responsibility of school-level professionals in most countries. At this level, summative assessment typically serves for school-internal purposes such as keeping records and giving reports of progress to students, parents and other teachers. Summative assessment for school-external purposes such as selection and certification tends to become more important as students progress to the higher levels of school education (Nusche, forthcoming).

Internal summative assessment

Internal summative assessment, implemented by teachers in the classroom, plays a key role across OECD countries. The majority of education systems have developed policy frameworks (national or state laws or regulations) that specify procedures for internal summative assessment, particularly in secondary education where summative assessment typically receives increased attention (Figure 4.2). These frameworks are generally developed at the central level[3] and they apply to all schools in the majority of education systems.[4]

While policy frameworks for internal summative assessment are commonplace in most countries, there are large variations regarding their level of detail and prescription. In Chile, for example, the policy framework sets very basic requirements, such as the number of assessments per year and requires schools to establish an internal protocol for assessment. In Finland, the requirements for internal summative assessment are included in the national core curriculum for general programmes and in the national qualification framework for vocational programmes. In Ireland, at ISCED levels 2 and 3, subject-specific assessment requirements are provided within the subject syllabi. In Poland, the framework is very general and leaves much autonomy for schools to set up their own assessment rules.

Figure 4.2 Existence of central frameworks for internal summative assessment at ISCED levels 1-3 (2012)



Source: Derived from information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability across countries.

A number of education systems do not have formal frameworks for internal summative assessment at any level of education, but certain basic requirements for summative assessment are typically set in the legislation, curriculum or regulations. For example, in the Flemish Community of Belgium, each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of central attainment targets and developmental objectives, and the Inspectorate may ask schools to present their policy. In the French Community of Belgium, the policy framework is provided through the work of the Inspectorate and the possibilities for students to appeal summative assessment decisions made by the class council. In the Czech Republic, schools are legally required to set their own assessment systems, which are approved by the school board and controlled by the Inspectorate. In Iceland, there is a requirement for students to undergo summative assessment at the end of Year 10, but the curriculum and regulations are flexible regarding how this is implemented by schools and it may in fact take the form of a formative assessment. In Norway, the Education Act states that students shall obtain summative achievement marks at the end of each year level in secondary education.

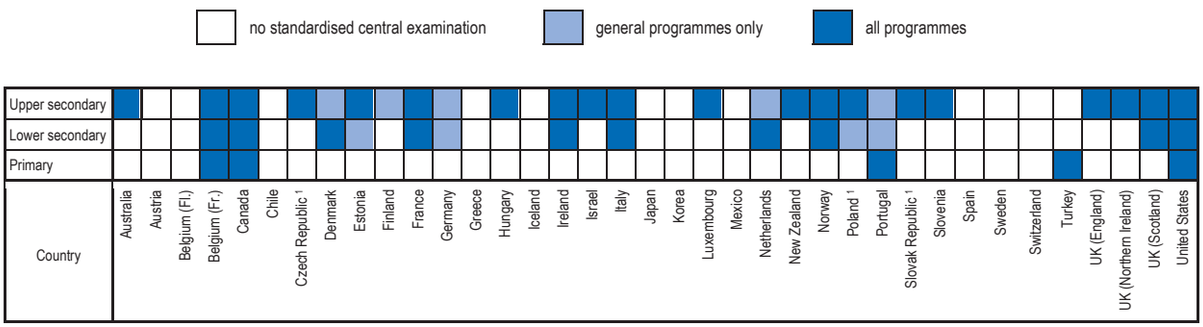
Regarding the primary education sector, it is important to note that some countries, such as Denmark and Norway, have an explicit policy to avoid summative marking and reporting in the first years of education, which explains the absence of a summative assessment framework at this level (Nusche et al., 2011a; Shewbridge et al., 2011).

Central examinations

In addition to internal summative assessments, many OECD education systems use central examinations, i.e. standardised assessments that are developed at the central level and have a formal consequence for students (e.g. influence on a student's eligibility to progress to higher levels of education) to measure student performance. Such examinations are rarely used in the early years of schooling and become more widespread at the higher levels. Of 37 OECD education systems for which information was available, only five used central examinations at the primary level (the French Community of Belgium, Canada, Portugal, Turkey and the United States), versus 14 education systems at the lower secondary level and 25 at the upper secondary level (Figure 4.3) (OECD, 2012a, complemented with information collected from countries participating in the OECD Review). In addition, standardised examinations offered to schools by private providers play an important role in some countries. In the Netherlands, for example,

85% of primary schools use the school leavers test developed by the Central Institute for Test Development (Cito), which provides information regarding the school type most suitable for each student in the next phase of education.

Figure 4.3 Existence of standardised central examinations at ISCED levels 1-3 (2012)



Notes: (1) Excludes ISCED 3C programmes at the upper secondary level.

Source: OECD (2012a), *Education at a Glance 2012*, complemented with information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

Assessment for formative purposes

As illustrated above, formative assessment, or "assessment *for* learning", aims to deepen and shape subsequent learning rather than making a judgement about past performance (Black and Wiliam, 1998). It is essentially a pedagogical approach consisting of frequent, interactive checks of student understanding to identify learning needs, provide feedback to students and adapt teaching strategies (OECD, 2005a). It is embedded in the normal day-to-day teaching and learning process and may include activities such as classroom interactions, questioning and feedback (Looney, 2011a). The use of assessment information is key to the concept of formative assessment: to be considered formative, assessment evidence must be acted upon in subsequent classroom teaching and learning. Students also participate actively in the process through self- and peer-assessment. In some recent formulations, the active participation of students in the process has given rise to the term *assessment as learning*, which focuses on students reflecting on and monitoring their own progress to inform future learning (Earl, 2003). The intentions for designing and implementing formative assessment strategies include:

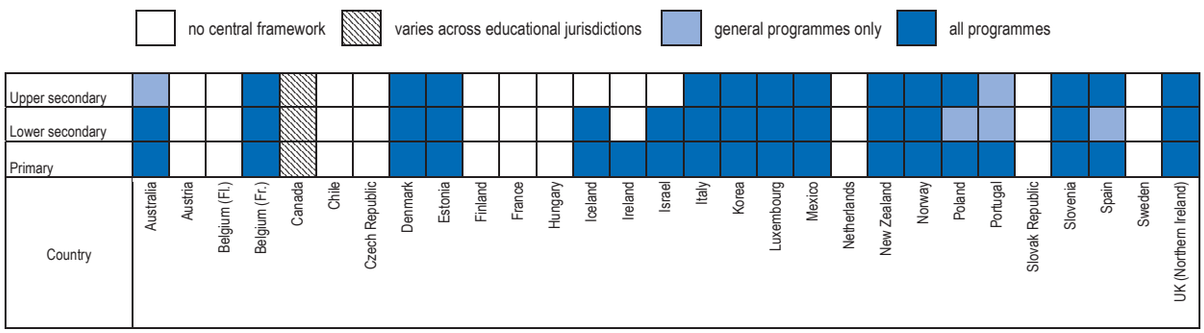
- **To provide timely feedback to students**, which they can integrate into their learning process. Several studies indicate that feedback is most effective when it is timely, is tied to criteria regarding expectations, and includes specific suggestions for how to improve future performance and meet learning goals (Wiliam, 2006; Hattie and Timperley, 2007; Swaffield, 2008).
- **To help students to feel safe to take risks and make mistakes** in the classroom. Students are thus more likely to reveal what they do and do not understand and are able to learn more effectively (Looney, 2011a).
- **To diagnose student learning needs and differentiate teaching accordingly**. In order to develop an appropriate teaching intervention, teachers need to assess students' learning needs and explore a range of potential causes of learning difficulties (Looney, 2011a).

- **To actively engage students in their own learning processes** so as to develop higher-order thinking skills and skills for "learning to learn", and to allow students and teachers to engage in conscious reflection on the learning process (Earl, 2003).

Internal formative assessment

Given the widely reported benefits of formative assessment for the improvement of teaching and learning, many OECD education systems have developed policy frameworks (national or state laws or regulations) to promote and support formative assessment practice in the classroom (Figure 4.4). Where these frameworks exist, they tend to be developed at the central (national or state) level and apply to all schools.[5] The existing frameworks generally include a requirement for schools to implement formative assessment in the classroom. In Australia (ISCED 2 and 3 only), Korea and Mexico, they also include a requirement for formative assessment to be part of initial teacher education programmes. In Korea, there is also a requirement for teachers to undertake professional development in this area. In Estonia, it is mandatory for schools to report on their strategies to promote formative assessment. In Spain, the regulations are most extensive including a requirement for schools to implement student formative assessment and to report on their strategies to promote student formative assessment, as well as for student formative assessment to be part of initial teacher education programmes and for teachers to undertake professional development in this area.

Figure 4.4 Existence of central frameworks for formative assessment at ISCED levels 1-3 (2012)



Source: Derived from information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

In some education systems, while formative assessment is not inscribed in national or state education law, it is promoted through other documents. In the Flemish Community of Belgium, primary schools are required to monitor the progress of every student and report observations to parents, but there are no specific regulations regarding the procedures for doing so. In Hungary, elements of formative assessment such as verbal assessment and differentiated assessment methods are included in legal regulations and the national core curriculum. In Ireland, the National Council for Curriculum and Assessment (NCCA) has issued guidelines to secondary schools that emphasise the value and use of formative assessment and while they are not in the form of regulations, they will be a key part of assessment requirements for the new ISCED 2 curriculum to be introduced on a phased basis from 2014. In Finland, the national core curricula for all ISCED levels mention that teachers should observe students' progress. In the Netherlands, draft laws are currently being prepared to set a requirement for schools to use formative assessment systems for

results-based work in schools. A number of countries, including Austria, Chile, the Czech Republic, France, Hungary, the Slovak Republic and Sweden do not have specific central regulations or documents promoting formative assessment.

It should also be noted that while existing policy frameworks signal the high level of attention given to formative assessment at the policy level, little information is available regarding the effective and systematic implementation of formative assessment across schools and classrooms. In a number of education systems participating in the OECD Review, the understanding and development of formative assessment appears to be still at an early stage of development. To ensure that policy commitments to formative assessment are matched with actual developments in the classroom, sustained investment in teachers' understanding and capacities regarding formative assessment is necessary. Box 4.3 provides some examples.

Box 4.3 Matching a commitment to formative assessment with concrete support for teachers

In **Canada**, many school districts offer professional development opportunities for teachers to improve their skills and knowledge of assessment/evaluation mechanisms. For example, over the past two years in particular, there has been a strong emphasis on Assessment for Learning practices in Nova Scotia schools. To that end, there was a provincial assessment summit in 2009 and several Boards then hosted their own Assessment Summits in 2010. The South Shore Regional School Board in Nova Scotia hosted a two day event in September 2010. As well, Assessment for Learning has been a Board priority in its Educational Business Plan and it remains so today. A website on assessment has been designed for teachers providing a multi-media workshop on the full scope of assessment knowledge, skills and applications (*http://web.ssrsb.ca/assessment/*).

In **Ireland**, the National Council for Curriculum and Assessment (NCCA) has contributed to the development of expertise in formative assessment through its curriculum development projects with schools. As part of its work with groups of teachers in its Primary School Network, the NCCA explores how formative assessment approaches can be implemented in Irish classrooms. The NCCA has also designed materials to support teachers and schools in expanding their assessment toolkit. Its Assessment for Learning website includes multi-media support and materials such as classroom video footage and samples of children's work with teacher commentary. There are also reflection tools and checklists to support individual teachers and whole school staffs in reviewing current assessment practice (*http://action.ncca.ie/primary.aspx*).

In **Norway**, a statutory requirement has been introduced for schools to implement assessment for learning. To support teachers in fulfilling the requirements for formative assessment, the Directorate for Education and Training has created a website on assessment for learning providing a range of materials and tools including questions for reflection, films, assessment tools and literature, and also examples of different ways to document formative assessment practice. At the same time, there has been a developing awareness that teachers have not traditionally received training in formative assessment and that there was very little expertise available nationally for school leaders to draw on to provide support. To address this, the Ministry of Education and Research and the Directorate for Education and Training in Norway identified formative assessment as a priority area for education policy and professional development and launched a range of support programmes and learning networks at the regional, local and school level. For example, the Assessment for Learning programme (2010-14) is organised in learning networks at the local and regional level, where practitioners can exchange experience and create spaces for common reflection on effective practice. Participating municipalities and counties employ a formative assessment contact person to assist in running the project locally. These contact persons attend Assessment for Learning workshops run by the Directorate. The programme also provides online resources including tools and videos on how to enact effective formative assessment in the classroom.

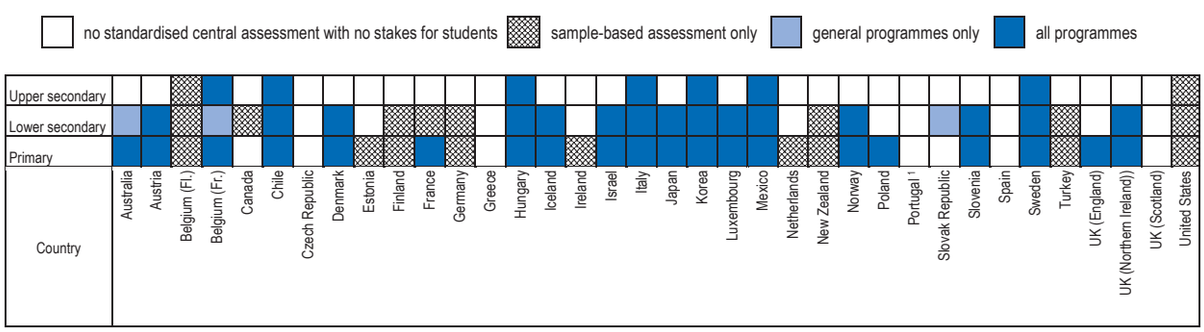
Sources: Fournier and Mildon (forthcoming); Irish Department of Education and Skills (2012); Nusche et al. (2011a).

Central assessments for formative use

Information from standardised assessments may also be used formatively to identify learning needs and adjust teaching strategies. Several education systems have developed standardised central assessments that do not have stakes for students. In some cases, these assessments were developed with the explicit purpose of providing formative or diagnostic information to teachers. In other cases, they are primarily aimed at producing aggregated performance information to monitor education system performance. The features of central assessments for system monitoring will be discussed primarily in Chapter 8 on system evaluation, but references will be provided in this chapter, because the results of such assessments may also be fed back to schools and be used by teachers in a formative way. Even where the central assessments test only a sample of students, the results may be provided to the participating schools.

Figure 4.5 below provides an overview of central assessments with no stakes for students across education systems. Clearly, such central assessments are most common at the primary (29 education systems) and lower secondary level (27 education systems), and less common at the upper secondary level (9 education systems). In most education systems, they assess the full cohort of students in the relevant years. In a number of education systems, these central assessments are sample-based only (Flemish Community of Belgium, Canada, Estonia, Finland, Germany, Ireland, the Netherlands, New Zealand, Turkey and the United States). France also has sample-based assessments at ISCED levels 1 and 2, in addition to full-cohort central assessments at ISCED level 1. England in the United Kingdom also has a sample-based assessment of science at ISCED level 1, in addition to full-cohort assessments of English and mathematics. In Ireland, all schools at ISCED level 1 are required to implement annual standardised assessments developed by private providers and a sample of primary schools participates in national assessments on a 5-year cycle.

Figure 4.5 Existence of standardised central assessments with no stakes for students at ISCED levels 1-3 (2012)



Note: (1) Before 2012/13 there were national assessments in Portuguese and mathematics.

Source: OECD (2012a), *Education at a Glance 2012*, complemented with information supplied by countries participating in the OECD Review. The figure should be interpreted as providing broad indications only, and not strict comparability.

Using large-scale assessments for formative purposes

Low stakes central assessments provide external signposts for teachers and students by indicating the learning goals that are expected nationally and can offer interesting pedagogical tools for teachers. In several countries where OECD Country Reviews took place, teachers were positive about such formative assessments and saw them as a tool to help them decide what they should be focusing on in their improvement plans to support

individual students. Positive effects of using student results from large-scale assessments to inform teaching may include: greater differentiation of instruction, greater collaboration among colleagues, an increased sense of efficacy and improved identification of students' learning needs (van Barneveld, 2008).

At the same time, in many settings there appear to be barriers to the effective use of large-scale assessments in the classroom. An important concern is that teachers often receive the results from such assessments too late to use them pedagogically. In fact, assessment data appear to have the most impact on student achievement when delivered in timely manner (Wiliam et al., 2004). Data from large-scale assessments, however, are sometimes available to teachers several weeks to months following the actual test day.

Moreover, data gathered in large-scale assessments are often not at the level of detail needed to diagnose individual student needs. Large-scale assessments typically include too few items on each topic for teachers to understand where students need help and what strategies might be useful in supporting them (Rothman, 2010). Especially for assessments that are also used for national monitoring, the scores are typically tied to broad proficiency categories, such as: below basic, basic, proficient, advanced (McGehee and Griffith, 2001). It has been argued that these categories are too broad to provide any kind of diagnostic information necessary for profiling individual student needs. In a study on the relationship between performance on a standards-based reading assessment and performance on a battery of diagnostic assessments, Rupp and Lesaux (2006) found that the standards-based assessment provided only weak diagnostic information, and masked significant heterogeneity in the causes of poor performance.

While large-scale standardised assessments can be useful to provide some initial clues about areas that need attention, other additional diagnostic assessments are needed to identify the causes of poor performance and develop an appropriate instructional intervention. Overall, while there is some evidence that data from large-scale assessments are being used successfully to identify students' strengths and weaknesses, to change regular classroom practice or to make decisions about resource allocation (Anderson et al., 2004; Shepard and Cutts-Dougherty, 1991), they need to be embedded in broader, more comprehensive assessment systems that include a range of summative and formative assessments, curriculum tasks, instructional tools, and professional development that helps teachers understand which assessment information is most appropriate for a particular purpose.

Balancing formative and summative assessment purposes

Finding a balance between formative and summative assessment is a challenge shared by many education systems. It is made more complex by the wide range of understandings of the meaning of the term "formative assessment", and the difficulty of managing the tensions between a stated commitment to formative assessment on the one hand, and public, parental and political pressure for accountability in the form of scores and rankings on the other (Harlen and James, 1997; Newton, 2007). This pressure for summative scores, and a conflation of formative and summative purposes in education policy documents sometimes results in confusion that in some cases may have hindered sound assessment practice, especially in the development of formative assessment (Harlen and James, 1997; Newton 2007). While the attention to results and data is a positive feature of education systems, an over-emphasis on these may have a negative impact, and undermine the formative role of teachers and assessment often so highly valued in policy goals.

As discussed above, there are cases in which there are misunderstandings and misinterpretations by teachers of the meaning and intentions behind formative assessment. In many education systems, teachers have long held the main responsibility for classroom assessment and see assessment as an important professional responsibility. At the same time, it is often not well understood that assessment for learning requires a major shift in mindset for teachers, as well as fundamental changes *vis-à-vis* traditional classroom assessment practices. While continuous classroom assessment, done by teachers on a regular basis, can include both summative and formative assessment, these labels represent fundamentally different purposes. Formative assessment is the process of identifying aspects of learning as it is developing, so that learning itself can be enhanced. It needs to be separate from the process of rating and ranking performance.

The challenge is to ensure that teachers move beyond surface techniques for formative assessment (such as "summative assessment done more often" or feedback that is unspecific or ego-involving) in order to adopt effective assessment for learning approaches. To have the greatest impact, feedback needs to provide information not only on how the learner has done, but also on the specific steps needed to progress further. It needs to be timely, detailed and specific (Wiliam, 2006; Hattie and Timperley, 2007; Swaffield, 2008). This requires ongoing professional learning opportunities and support for teachers, as well as suitable strategies to integrate classroom-based formative assessment within a broader framework for student assessment.

Developing coherent assessment frameworks, where each assessment is fit for purpose

In many education systems participating in the OECD Review, the introduction of central examinations and assessments is relatively recent and in some instances the communication about their purposes has not been sufficiently clear. If a specific assessment was designed for a particular purpose, its use for a different purpose is not ideal because it is likely that inferences made based on the test result will not be accurate, valid or useful for other purposes. For example, if the purpose of an assessment is diagnostic, then the test needs to be designed to provide very fine-grained information which allows uncovering specific difficulties and misconceptions of individual students. On the other hand, if the purpose of assessment is to compare schools or regions, the test needs to be designed to provide highly reliable summative scores in broad comparable categories (such as the number of students meeting or not meeting standards). It also needs to be complemented by broader information about the school's context, characteristics and processes (see Chapter 6).

There are risks in using a single test for too many purposes, in particular where the information ideally required in each case is not the same (Morris, 2011). In some countries participating in the OECD Review, national assessments have been introduced as low stakes and formative assessments without impact on individual students, teachers or schools. Subsequently, however, additional uses were added to the test, for example they were used to appraise teachers or to monitor school performance. Such shifting purposes may undermine the credibility of the assessment and jeopardise the constructive use of the test's results.

Since assessments need to be designed in line with their specific purpose, it is important for policy makers and assessment developers to be clear from the beginning about the main purpose for which assessment results will be used. According to Newton (2007), where multiple purposes are intended to be achieved, it is essential to clarify the *primary purpose* of

the assessment and to prioritise explicitly among purposes. The primary purpose and, to the extent possible, other high priority purposes should then determine assessment design characteristics to ensure the validity of results for intended purposes.

Newton (2007) suggests that the use of results for more than one purpose is possible if the purposes are not logically incompatible and if there is clarity about the type of evidence the assessment can provide. For example, an assessment that is designed primarily to help education policy makers monitor the education system should also be of value to those who participate in the assessment, and the results should be fed back to them along with illustrations of the types of inferences they can and cannot draw from the results. However, an assessment that is designed primarily for diagnostic and formative purposes should not be used for summative or accountability purposes because this may undermine its primary function (Linn, 2000).

Box 4.4 Defining and communicating the purposes of assessment

In **Canada**, the *Principles for Fair Student Assessment Practices for Education in Canada* outline key elements for assessment practice that have served as foundations for teacher handbooks, board polices and departments of education policy documents on assessment and test development in all Canadian jurisdictions. The Principles were developed in response to what was perceived as assessment practices not deemed appropriate for Canadians students. These principles and guidelines intended for both assessment practitioners and policy makers identify the issues to be taken into account in order that assessment exercises to be deemed fair and equitable. The text acts both as a set of parameters and a handbook for assessment. The first part deals with developing and choosing methods for assessment, collecting assessment information, judging and scoring student performance, summarising and interpreting results, and reporting assessment findings. It is directed towards practising teachers and the application of assessment modes in the classroom setting. The second part is aimed at developers of external assessments such as jurisdictional ministry/department personnel, school boards/districts, and commercial test developers. It includes sections on developing and selecting methods for assessment, collecting and interpreting assessment information, informing students being assessed, and implementing mandated assessment programs (for more information, see: *www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf*).

In **Denmark**, official information on the national tests produced by the former School Agency clearly repeats the message that the national tests only measure a discrete area of student knowledge and skills and teachers should use a range of other tests to gauge student progress. For example, it is stressed that the Danish test only measures students' proficiency in reading and a wide range of key knowledge and skills in Danish (e.g. spelling, grammar, punctuation, cultural understanding, literary knowledge, ability to express oneself) is not tested. Educators are aware that the tests provide only a snapshot of students' achievement levels in select learning targets and subjects (Wandall, 2010).

The **New Zealand** Ministry of Education *Position Paper on Assessment* (2010) provides a formal statement of its vision for assessment. It describes what the assessment landscape should look like if assessment is to be used effectively to promote system-wide improvement within, and across, all layers of the schooling system. The paper places assessment firmly at the heart of effective teaching and learning. The key principles highlighted and explained in the paper are: the student is at the centre; the curriculum underpins assessment; building assessment capability is crucial to achieving improvement; an assessment capable system is an accountable system; a range of evidence drawn from multiple sources potentially enables a more accurate response; effective assessment is reliant on quality interactions and relationships. To support effective assessment practice at the school level, the Ministry of Education is also currently conducting an exercise which maps existing student assessment tools. The purpose is to align some of the assessment tools to the National Standards and provide an Assessment Resource Map to help school professionals select the appropriate assessment tool to fit their purpose.

Sources: Fournier and Mildon (forthcoming); Shewbridge et al. (2011); Nusche et al. (2012).

In this context, it is important that all stakeholders have a broad understanding of assessment and of the need to combine a range of different assessment information in order to come to good judgements and valid interpretations on student learning and progress (Absolum et al., 2009). Teachers need skills to interpret standardised assessment results, to understand where further diagnostic testing of some students is necessary and to identify areas were teaching strategies may need adjustment to meet student needs.

A key governance challenge for countries is to develop a clear vision and strategy for assessment where different approaches developed nationally and locally each serve a clearly defined purposes and the format of the assessment is aligned to these particular purposes. In New Zealand, for example, the Ministry of Education has published a position paper on assessment in 2010, outlining the underlying principles of assessment. This has been complemented by a mapping exercise to clarify the distinct aims of existing standardised assessment tools that are at teachers' disposal (Box 4.4). Also, clear communication is vital to ensure that assessment results are used in an effective way. In particular, it is important to establish safeguards to avoid an over-emphasis on standardised assessment results. One such safeguard is to communicate clearly about the kinds of evidence that different types of assessment can and cannot provide (for examples from Canada and Denmark, see Box 4.4).

Responsibilities for student assessment

This sub-section provides an overview of the different actors involved in student assessment and the ways in which the different levels of governance interact. Responsibilities for student assessment are typically shared between different agencies and levels of the education system. In most countries, teachers have the main responsibility for continuous formative and summative assessment in the classroom, whereas regional, state or national agencies tend to be in charge of developing and marking standardised assessments that are used at key stages of education.

Internal assessment

In most OECD countries, schools benefit from considerable autonomy in the organisation of internal student assessments. School leaders, together with teachers, and sometimes in co-operation with school governing boards or education authorities, are typically in charge of establishing school policies for student assessment. Across the OECD, in PISA 2009, 66% of 15-year-old students were in schools whose principals reported that the school alone had the main responsibility for establishing student assessment policies, and 23% of students were in schools where the school together with the regional and/or national education authority had considerable responsibility for student assessment policies (OECD, 2010a).

While schools tend to have considerable freedom in establishing their own assessment policies, certain basic requirements are generally set in central policy frameworks (see above). The frameworks for internal summative assessment are centrally defined in most education systems, but in fact, different levels of education are involved in ensuring compliance with these frameworks. Across OECD education systems, this task may be attributed to the school leaders and teachers (e.g. Chile, Estonia, Finland, France, Hungary[6], Israel, Poland, Slovak Republic, Slovenia), the subject committees (Korea), the schools boards (e.g. Czech Republic, Iceland, Ireland and New Zealand), the school organising bodies (Hungary, Netherlands), the local education authorities (Norway), the central, state or provincial authorities (Australia, Austria, Canada, Luxembourg, Mexico,

New Zealand) or the Inspectorates (Flemish and French Communities of Belgium, Czech Republic, Netherlands) (Tables 4.A1.1, 4.A2.1a, 4.A2.1b, Annexes 4.A1 and 4.A2).

In their national background reports prepared for this Review, the majority of education systems reported about long-standing traditions of teacher-developed assessment and the historically important role of teachers' professional judgements in assessment. Teachers are generally expected to take responsibility for different functions of assessment including diagnostic, formative and summative. While teachers tend to have the exclusive responsibility for summative assessment in primary education, their assessment approaches are typically complemented by regionally or nationally implemented standardised examinations at the higher levels of education (Figure 4.3). The distribution of responsibilities tends to be organised in a way that teachers assess and report on student performance in relation to the full range of curriculum goals, while standardised examinations and assessments assess a particular subset of learning goals in specific year levels.

Regional and central examinations

Tables 4.A1.4, 4.A2.4a and 4.A2.4b (Annexes 4.A1 and 4.A2) provide detailed information about the groups involved in national student examinations that have a formal consequence for individual students. Student examinations are considered "standardised" if they are designed to ensure that the questions, conditions for administering, marking procedures, and interpretations are consistent and comparable among students (Popham, 1991). To ensure these conditions, examinations are often designed and marked at a central level. At the same time, many countries have hybrid forms where assessments are centrally developed but locally administrated and/or marked. In this case, countries tend to use guidance materials and moderation to ensure the reliability of local marking.

At the lower secondary level, the central education authorities have full responsibility for developing national examinations in the French Community of Belgium, Denmark, France, Italy and Norway, while they share responsibility for this task with a central agency for assessment in Australia, Estonia and Mexico. In some countries, including Ireland, the Netherlands and Poland, central/regional agencies for assessment hold the main responsibility for the development of external examinations. While the examinations are centrally developed in all education systems, school level examiners play a key role in the marking process in several education systems including the French Community of Belgium, Denmark, Estonia, France, Italy and the Netherlands. Where high stakes examinations are marked locally by teachers, moderation arrangements are typically in place to ensure the reliability of marking, for example through the involvement of a second marker in addition to the students' own teachers (more on this below).

At the upper secondary level, the central education authorities have full responsibility for developing national examinations in the French Community of Belgium, Denmark (general programmes only), France, Hungary (general programmes only), Israel, Italy, Luxembourg, New Zealand and Norway. They share responsibility for this task with a central agency or institute for assessment in Australia, the Czech Republic, Estonia and Mexico. Central or regional agencies for assessment hold responsibility for developing examinations in general education programmes in Finland, Ireland, the Netherlands and Slovenia, and for developing examinations in both general and vocational programmes in the Slovak Republic and Poland. In vocational programmes, the schools themselves play a key part in developing examinations in Denmark and Slovenia.

The examinations in upper secondary education are centrally marked by the education authorities and/or a central agency in most education systems. At the same time, school level examiners play a key role in the marking process in a range of education systems including the French Community of Belgium, the Czech Republic, Denmark, France, Italy, Luxembourg, the Netherlands and Slovenia (vocational programmes only). In Hungary, the marking is undertaken at the school level for the "normal"-level examination and at the central level for the advanced-level examination. In New Zealand, internally assessed subjects are marked in the school whereas externally assessed papers are marked by a central agency (the New Zealand Qualifications Authority). In Norway, centrally-given written examinations are externally marked whereas locally-given oral examinations are marked together by the subject teacher and an external examiner. Examinations in vocational programmes are marked by the local education authorities.

Assessments developed by other providers

Private providers play an important role in test development in several education systems. No internationally comparable information is available regarding the importance of private testing companies across countries. From the OECD Country Reviews, it appears that the use of private tests by schools is commonplace in most countries. In some education systems, such as Ireland (ISCED 1 at present and ISCED 2 from 2014), schools are required to choose from among certain standardised tests developed by private providers and to report the results to their school boards and the educational administration.

In the Netherlands, schools are required to report on students' learning results in the final phase of primary education in a way that clarifies the extent to which students have reached the minimal achievement levels for primary education. While schools are free to use different assessment instruments for this purpose, the vast majority of primary schools use to this end the school leavers test developed by the Central Institute for Test Development (Cito). The results from this test also provide information on the school type which is most suitable for each student in the next phase of education.

In Northern Ireland in the United Kingdom, upper secondary schools are expected to choose standardised central examinations from a range of Awarding Organisations. The relevant Awarding Organisations are responsible for marking the external assessment and where internal assessment is used, they also moderate teachers' marking.

In other countries, such as the Czech Republic, private companies offer testing services that schools can choose to buy to support their regular assessment practice. In several education systems, such as the Flemish Community of Belgium and the Slovak Republic, classroom-based assessment is heavily influenced by textbooks, many of which contain tests.

The potential influence of private assessment companies on teacher assessment practices is substantial. Commercial tests available to schools typically aim to provide summative data of the students' level of knowledge in different subjects. In some countries, private providers also offer tests explicitly designed for formative and didactic purposes (for an example from the Netherlands, see Box 4.5). Schools receive feedback of results providing information on areas for individual students to improve. Certain commercial tests may enjoy direct support by local education authorities and consequently may allow the feedback to schools of comparative data, as schools throughout the municipality use these tests. In the Flemish Community of Belgium, school umbrella organisations make tests available to schools belonging to their network. As such, these tests provide information on each student's competencies, but also a

benchmark for schools to compare to other schools within the network. The use of additional externally developed tests can provide schools with useful information as part of their wider student assessment systems.

Box 4.5 Monitoring student learning in the Netherlands

Since the mid-eighties primary schools started to make use of a pupil monitoring system, the LVS (*Leerling Volg Systeem)* developed by the Central Institute for Test Development (Cito). Later on pupil monitoring systems were also implemented in secondary schools and currently every secondary school has a pupil monitoring system. The Cito pupil monitoring system (LVS) for primary education is a consistent set of nationally standardised tests for longitudinal assessment of a pupil's achievement throughout primary education, as well as a system for manual or automated registration of pupil progress. The LVS covers language, (including decoding and reading comprehension), arithmetic, world orientation (geography, history, biology), social-emotional development, English, science and technology. It is purchased by schools at their own cost and initiative. The primary objective of the LVS is the formative assessment of student achievement and individual students' mastery of key subject matter areas in relation to their year level. Item Response Theory is used to vertically equate students' scores in the LVS tests, which allow for a calculation of student growth trajectories in primary school. Since 2003, the LVS also contains computer-based tests, some of which are adaptive. The following presentation formats are made available on the basis of the LVS:

- The *pupil report,* which is a graph in which the pupil's progress is visible throughout the years. Data available in the national surveys are used as a frame of reference, based on percentiles, so that the position of an individual pupil with regards to five reference groups (25% highest scoring pupils, just above average, just below average, far below average, and the 10% lowest scoring pupils) is immediately visible from the corresponding graph.
- For children with special education needs, and who visit special education schools, an *alternative pupil report* is made available. This report also shows at what level a pupil is functioning and how to interpret the results of the pupil compared to children of the same age who attend mainstream primary education.
- In the so called *group survey* the results of all the pupils from a group over a number of years are presented in a table. For each pupil the scale of ability score at the successive measuring moments is shown along with the level score.

Source: Scheerens et al. (2012).

However, in some countries participating in the OECD Review, there are a number of concerns related to the regular use of private tests in the classroom. Often, the most widely used commercial tests contain mainly multiple-choice items and closed-format short answer questions, which are best suited to assess knowledge-based elements of the curriculum. In some cases, commercial tests were being perceived as practice tests for national assessments, hence reinforcing a focus on assessing only a limited subset of learning objectives that can be computer-scored. It was not always clear to the OECD review teams that the tests offered to schools were closely aligned with national curricula. Also, teachers who use commercial tests irrespective of the national and local education goals and without eliminating non-relevant content may present their students with too much content at the expense of essential national or local learning objectives. It is important that independent information about the quality and relevance of private tests is made available to teachers including to what extent they offer useful feedback on student progress against the national learning objectives. Ideally, there should be an accreditation process to validate the use of such tests as reflecting national student learning objectives.

Balancing external assessments and teacher-based assessments

Many countries rely on a mix of external and internal assessment, but finding the right balance between the two approaches may be challenging.

The major advantage of external standardised assessment is its high reliability. It ensures that all students are assessed on the same tasks and that their results are measured by the same standards. Standardised external assessment is usually conducted in supervised conditions which ensure that what is assessed is the students' own work (Crooks, 2004). It is marked by a machine or by external assessors and the marking criteria are standardised, so that a high degree of reliability is given. The marking is expected to be free of bias or discrimination, as the assessors do not know the students whose work they are reviewing. The results are made as objective as possible so that they are, within a year, comparable among students, regardless where they go to school (Rosenkvist, 2010). Externally defined assessments can clearly indicate the standards that are expected nationally of all students, so that they can steer their learning in that direction (Elshout-Mohr et al., 2002).

However, external assessment is often criticised for having lower validity than teacher-based assessment. It tends to be in the form of a written test under supervised conditions, so that only a limited range of curriculum goals can be covered. Also, external assessment typically takes place on very few occasions and thus gives limited information about students' competencies due to the normal daily variations in performance. It can also have detrimental effects on teaching and learning, as teachers may end up focussing on test-taking skills, especially when high stakes are attached to the test results (Box 4.1). The high stakes that are often attached to a single external examination can cause stress or test-anxiety among students resulting in their achievements being reduced on the examination day (Crooks, 2004).

Internal assessment also has its advantages and drawbacks. Due to its continuous nature, teacher-based assessment allows for important achievements to be measured that are more difficult to capture in an external examination, such as extended projects, practical assignments or oral work. Internal assessment thus has a higher potential for the full range of curriculum goals to be covered (Crooks, 2004; Harlen, 2007). As internal assessment is embedded in the regular coursework and spread throughout the course, it is also more authentic than a test-based external examination, providing more opportunities for students to show what they know and are able to do in normal conditions. Especially if particular outcomes are assessed several times, atypical performances as well as achievement trends can be identified (Crooks, 2004).

However, it is important to note that the validity of teacher-based assessment depends to a large extent on the assessment opportunities provided by individual teachers. It is difficult to ensure that all teachers indeed use the potential of internal assessment to cover the full range of goals specified in the curriculum (Harlen, 2007). Several reviews of research on teacher-based assessment note that teacher-made assessments are often no more diverse or innovative than external assessments, encouraging rote learning and recall of fragmented knowledge rather than critical thinking and deeper learning (e.g. Crooks, 1988; Black, 1993; Black and Wiliam, 1998; Harlen, 2007). For internal assessment to work well, it is essential to ensure that teachers receive adequate training to develop their assessment skills (see section on "Capacity").

Internal assessment is often perceived as being less reliable than external assessment. Assessment items and marking standards may vary widely between teachers and schools,

so that the results of internal assessment will lack external confidence and cannot be compared across schools. It cannot always be verified that what is assessed is indeed the student's own work, as some tasks for assessment (e.g. homework and project work) may take place outside the classroom with little supervision (Crooks, 2004). Several studies also report that there is a risk of conscious or unconscious bias in teacher-based assessment, i.e. that teachers may give more help to some students, or in their marking may give inappropriate weight to prior knowledge and expectations of particular students (Crooks, 2004; Harlen, 2007).

Where teachers are responsible for summative assessment, there is also a risk that teachers become subject of parental pressure to lower assessment standards and provide higher marks for their children (Bishop, 2006). Figlio and Lucas (2004) find that parents do not perceive tougher teachers to be better teachers and tend to prefer high marks over high standards. This could lead to potential distortions of results due to parental pressure. With internal summative assessment, the teacher also acquires a double role of teacher and assessor. Rather than strengthening the teacher-student relationship, this may in fact result in a distancing between the student and the teacher (Bishop, 2006). Indeed, students may refrain from asking questions fearing that this could be interpreted as a sign of slow progress and low achievement (Somerset, 1996). The teachers' role in internal summative assessment may thus negatively impact on the effectiveness of their formative assessment approaches.

Research describes several ways to address potential bias in teachers' assessment and increase the reliability of the assessment. There is evidence that the reliability of teacher-based assessments can be improved by the use of scoring guides detailing descriptions of competency levels and providing examples of high performance (Harlen, 2004; 2005). There are also indications that teachers apply assessment criteria more accurately if they are clear about the goals to be achieved and especially if they have participated in the development of criteria (Hargreaves et al., 1996; Frederiksen and White, in EPPI, 2004). External benchmarks showing what is considered to be normal or adequate progress of students in particular marks and subjects are also helpful to help teachers make accurate judgements. In Sweden, for example, teachers are encouraged to compare the achievements of students in internal assessment to student results in national assessments and to use the national assessments as an external guidance and reference points (Nusche et al., 2011b). Finally, training for teachers, teacher collaboration in assessment and external moderation of teacher-based assessment can further enhance the reliability of internal assessments (more on this in the section on "Capacity").

Crooks (2004) suggests that a combination of teacher-based and external assessments would be most suitable to ensure maximum validity and reliability. Learning outcomes that can be readily assessed in external examinations should be covered this way, whereas more complex competencies should be assessed through continuous teacher-based assessment. Where teacher-based assessment is used for summative purposes, it is essential to pay attention to maximising reliability, by using scoring guides, negotiated scoring criteria, external benchmarks, training for teachers, multiple judgements and external moderation. It is also important to provide a range of nationally validated assessment tools that teachers can use to assess their students reliably when they see fit.

Reference points for student assessment

Clear and explicit expectations for student learning and assessment criteria are important to ensure the validity, reliability, transparency and usability of assessment (for

definitions, see section on "Analytical approach"). Information on expected learning outcomes and developments are typically expressed in national curricula, educational standards or learning progressions.

- **National curricula** typically describe overarching learning objectives for the education system and explain the underlying values and culture that should shape teaching and learning. Countries take different approaches to how they design curricula. While some describe the teaching content, methods, materials and assessment criteria to be applied in different subjects and year levels, others establish broad guidelines, leaving room for local authorities and schools to decide upon more specific goals, content and methods. A national curriculum typically covers all subjects and courses offered in school education, whereas a **syllabus** provides more detailed information regarding the expectations for an individual school subject.
- **Educational standards** refer to descriptions of what students should know (content standards) and be able to do (performance standards) at different stages of the learning process. In some countries, standards are only available for the core subjects, such as literacy and mathematics, whereas in other countries they exist for a broad range of subjects. The standards may be set out in a separate document, or may be embedded in the curriculum.
- **Learning progressions** describe the way students typically move through learning in different subject areas. They can provide a roadmap for teachers to identify the set of skills and knowledge students must master on the way to becoming competent in more complex curriculum outcomes. Such learning progressions may be described in the curriculum or a separate document.

Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2) provide an overview of the references used for student assessment across OECD education systems at different levels of education. The references used in lower secondary education are listed in Table 4.1 below. As can be seen from the table, the majority of education systems use central or state curriculum goals as the main reference for student assessment. Central standards are also frequently used, either as the main reference for assessment or in addition to national curriculum goals. Central standards are used for at least one assessment type (internal summative assessment, national examinations and/or national assessments with no stakes for students) in Australia, Austria, the Flemish and French Communities of Belgium, Chile, the Czech Republic, Estonia, France, Hungary, Luxembourg, the Slovak Republic, Poland and Spain. Learning progressions are less frequently used across education systems. They serve as references for internal summative assessment and central assessments in Northern Ireland in the United Kingdom, for central assessments in Australia and Norway and for both national assessments and examinations in Denmark. Poland has specific examination standards based on the national core curriculum that serve as references for central examinations and Canada uses its own assessment framework for the development of the Pan-Canadian Assessment Program. In Denmark, there is no national curriculum in compulsory education, but binding national objectives were introduced in 2003 to serve as a reference for internal summative assessment. In Sweden, national knowledge requirements serve as the main reference for central assessments.

Table 4.1 References used in student assessment in lower secondary education – ISCED 2 (2012)

	Internal summative assessment	Central examinations	Central assessment
Central curriculum goals	Australia (pre-voc and voc), Austria, Belgium (Fl.), Belgium (Fr.), Finland, France, Iceland, Italy, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Slovenia, Spain, Portugal	Estonia (general), France, Ireland, Italy, Netherlands (pre-voc and voc), Norway, United Kingdom (Northern Ireland), Portugal[1]	Iceland, Israel, Italy, Korea, Mexico, Netherlands (pre-voc and voc), New Zealand, Slovak Republic (general), Slovenia
Central standards	Poland (plus curriculum decided at the school level), Slovak Republic	Belgium (Fr.) (general)	Austria, Belgium (Fl.), Belgium (Fr.) (general), Luxembourg
Central curriculum goals and standards	Australia (general), Chile, Czech Republic, Estonia, Hungary		Chile, Czech Republic, France, Spain
Learning progressions	United Kingdom (Northern Ireland)	Denmark	Australia, Denmark, Norway, United Kingdom (Northern Ireland)
Specific examination/ assessment standards		Poland (examination standards based on core curriculum) (general)	Canada (Pan-Canadian Assessment Framework), Finland (marking guidelines), Hungary (National Assessment Framework)
Binding national objectives/national knowledge requirements	Denmark		Sweden

Note: (1) National standards will be used as a reference for central examinations from 2013/14.

Source: Derived from information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Providing clear goals and expectations for student learning to guide assessment

There are pronounced differences in the degree to which countries set system-wide expectations for student performance to guide teaching, learning and assessment across schools. While it is common for OECD countries to have system-wide curriculum frameworks that set objectives for student learning, the degree of prescription varies widely between countries.

Highly prescriptive central curricula can act as a legal trigger to promote evidence-based approaches and innovations regarding content and pedagogy and bring these within the reach of all schools and teachers (Elmore and Sykes, 1992; Westbury, 2007; in Kärkkäinen, 2012). But such curricula may not respond well to different local contexts and there may be a lack of ownership and commitment to change among teachers. On the other hand, highly decentralised curricula allow schools and teachers to experiment and develop curriculum innovations that are relevant to local contexts and may spread through horizontal networks of schools (Elmore and Sykes, 1992; Elmore, 1996; Darling-Hammond, 1998; Marsh and Willis, 2007; in Kärkkäinen, 2012).

Several education systems have adopted participatory approaches to developing curricula, where the national curriculum provides the core of overarching objectives whereas the more specific goals and curriculum content and assessment criteria are developed at the local and/or school level. This is the case, for example, in the Flemish Community of Belgium, the Czech Republic, Denmark, Finland, Norway and the Slovak Republic. Such curricular autonomy is intended to provide space for local interpretation and adaptation of goals. It is also expected to help strengthen local ownership of the teaching programme.

While it is important to keep curricula open so as to allow for teachers' professional judgements and innovations in the classroom, in some settings there are concerns about a lack of clarity regarding the specific goals to be achieved by all students in different subjects and year levels. There are often large variations in schools' capacity and expertise to implement effective local curricula and assessment approaches. This may lead to a lack of equivalence and fairness in educational opportunities for students across the country. In the absence of clear and specific system-wide objectives, teachers may find it difficult to develop concrete lesson plans, learning goals and assessment strategies that are in line with national expectations.

The introduction of more detailed national standards (or expectations or benchmarks or competence goals) for what should be taught, learned and assessed in schools has been debated and tried to varying extents in many countries over the last quarter century. Central standards are intended to provide consistency and coherence, especially in contexts where there is a high degree of local autonomy regarding the development of curricula, teaching programmes and assessment. Although it may appear straightforward to create statements of expected learning and levels of proficiency, experiences in different education systems have shown that it is not an easy task to identify clear and agreed standards and criteria (Looney, 2011b; Nusche et al., 2011a).

Research also reveals challenges in ensuring that the curriculum, standards, teaching and assessment are consistent (see also Chapter 3). The core logic of standards-based systems rests upon the alignment of these key elements. If the assessments do not well match the curriculum and the standards, then assessment results have little value in judging how well students are learning. This, in turn, will make it difficult to diagnose and respond to student or school needs. Hence, policy needs to give considerable attention to sound strategies that assess student performance in relation to the curriculum and the standards. In addition, teacher education and professional development also need to be aligned to overarching curriculum goals and standards. Where curricula and assessments have been reformed to focus on key competencies, it is important to ensure that teaching and learning approaches are changed accordingly to provide opportunities for students to indeed acquire these competencies.

Procedures

This second section describes the procedures and methodologies used for student assessment across countries. This includes the scope of assessment, i.e. the areas of learning that are covered by the assessment as well as the key procedural features of student assessment across countries, i.e. the mix of instruments used in specific student assessment systems; the format of assessments; and the use of ICT in assessment. It also reviews ways in which the design of assessments can enhance or threaten fairness and equity in education.

Aspects assessed

A comprehensive body of research has described the different dimensions of learning that might be captured in assessment. Traditionally, classifications of learning outcomes have been based on the distinction between cognitive and non-cognitive learning, but more recently the concept of competencies which encompasses both cognitive and non-cognitive learning has become widely accepted. Box 4.6 provides an overview of different types of learning outcomes that education systems may seek to achieve.

While there is a strong ambition across OECD countries to focus school systems increasingly on the development of complex competencies, the OECD Country Reviews found a concern across countries that assessment systems might be lagging behind such competency-based curricula. Both standardised and teacher-based assessment often remained more traditional and focussed on isolated knowledge and the application of basic skills. While no directly comparable information is available regarding the scope of student assessment across countries, information regarding the subjects assessed and the assessment instruments used provides some indications about the scope of learning that is typically captured in current approaches to student assessment.

Box 4.6 Classification of learning outcomes: A brief overview

Cognitive learning refers to the construction of thought processes. Most current classifications of cognitive learning are inspired by the "taxonomy of educational objectives" developed by Bloom in the 1950s. Bloom divided the cognitive domain into six categories of increasing complexity: knowledge, comprehension, application, analysis, synthesis and evaluation (Bloom, 1956). Bloom's taxonomy has since been further developed, modified and refined by many authors (e.g. Gagné, 1977; Klemp, 1977; Kolb, 1981; Eraut, 1990, Marzano, 2001). Most authors of the assessment literature have retained two major categories of cognitive learning: the *acquisition of knowledge* and the *development of skills* (Posner, 1992). The acquisition of knowledge involves the recognition or recall of ideas, materials and phenomena, whereas the development of skills involves more complex processes of analysis, synthesis and evaluation. Sometimes, within the "skills" category, an additional distinction is made between *basic skills*, such as the routine application of well-practiced knowledge and skills, and *higher-order skills,* such as problem solving and the transfer of existing knowledge and skills to new situations (Crooks, 1988).

Non-cognitive learning refers to psychosocial development and the development of attitudes and values. Psychosocial development can include self-development (e.g. self-esteem, identity development) as well as relationships with other people and institutions (e.g. interpersonal and intercultural skills). Attitudes and values are closely related. *Attitudes* can be defined as beliefs focused on a specific object whereas *values* refer to more generalised standards that transcend attitudes (Pascarella and Terenzini, 2005). Attitudinal and value outcomes of learning can include development of social responsibility, motivation for learning and understanding of diversity (Volkwein, 2003).

Competency development refers to a broader model of learning which comprises a complex integration of knowledge, skills, attitudes, values and action (Baartman et al., 2006). A competency can be defined as "the ability to meet demands or carry out a task successfully and consists of both cognitive and non-cognitive dimensions" (Rychen, 2004). Competencies are complex ability constructs that are applied in real-life contexts and have to be acquired by learning. The concept of competency is based on the idea that the traditional taxonomies of learning cannot adequately capture the nature of learning that unites different types of knowledge, skills and attitudes into real expertise (Ewell, 2005). In recent years, many OECD countries have reformed their education goals and/or curricula to emphasise the importance of "key competencies" or "21st century competencies". While the definition of such "key competencies" varies across countries, they typically include elements of learning to learn skills, information processing, communication, teamwork, critical analysis and creativity.

Source: Reproduced from Nusche (2008).

Subjects assessed

In all OECD countries, schools are expected to monitor their educational quality and assess the extent to which students acquire the knowledge, skills and competencies outlined in the curriculum or performance standards. Typically, students are assessed by their teachers in all subjects that are part of the curriculum. Assessments tend to be continuous throughout the year and are generally complemented by more high stakes teacher-based or external examinations at certain key stages in the school year or cycle.

While it is generally expected that all subjects are given some attention in teacher-based assessment, standardised assessments and examinations – where they exist – tend to focus on a few priority subjects. The subjects assessed in central examinations vary across education systems. As shown in Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2), the subjects most frequently assessed in OECD countries are the language of instruction and mathematics, at all levels of education. But in many education systems, a range of other subjects are also assessed.

Subjects covered in central examinations

Only a few OECD education systems have central examinations in place in primary education (Figure 4.3). The French Community of Belgium, Portugal and the United States assess all primary school students in mathematics and the language of instruction. In addition, both the French Community of Belgium and the United States assess students in natural sciences, and the French Community of Belgium also assesses students in social sciences. In Turkey, national examinations exist in primary education in Year levels 6, 7 and 8, but they are not compulsory and students are given a choice of subjects for examination (OECD, 2012a).

Fourteen OECD education systems implement central examinations in lower secondary education. Thirteen of them are depicted in Table 4.2 below (Canada is not included in Table 4.2 because the characteristics of examinations vary across provinces). As shown in the table, all education systems except Norway and Scotland in the United Kingdom have some examination subjects that are compulsory for all students at this level. The subjects most frequently examined in a compulsory manner are the language of instruction (11 systems) and mathematics (10 systems). Other frequently examined compulsory subjects include the sciences and modern foreign languages. In Scotland, there are no compulsory examination subjects, but students choose among a range of possible subjects. In four countries (Denmark, Estonia, Ireland and the Netherlands), students can choose among a range of optional subjects in addition to a number of compulsory subjects. Norway and Denmark use an approach where students are randomly selected to sit an examination in a given subject. In Norway, students in Year 10 are randomly sampled to sit a centrally administered written examination in one subject among Norwegian, mathematics and English. In Denmark, in addition to a range of compulsory and optional examination subjects, students are randomly sampled to sit examinations in social sciences and religion.

Twenty-five OECD education systems have central examinations in upper secondary education, and most of them have at least one compulsory examination subject that all students have to take (OECD, 2012a, complemented with information collected from countries participating in the OECD Review). The language of instruction is the most frequent compulsory examination subject, followed by mathematics and modern foreign languages. Examinations in upper secondary education are characterised by a wider range

of options for students. While in lower secondary education less than half of the countries have optional examination subjects for students, in upper secondary education this is the case for over two-thirds of the countries. As in lower secondary education, Denmark and Norway use a sampling approach, where students are randomly selected for examinations in most subjects (OECD, 2012a).

Table 4.2 Characteristics of central examinations in lower secondary education – ISCED 2 (2011-12)

Country	Programme type	Are central examinations compulsory?	Which year levels are assessed?	What subjects are assessed?		
				All students	Students choose	Sample
Belgium (Fr.)	General only	Yes, for public schools only (mandated at the Community level)	8	M, L		
Denmark	All programmes	Yes, for public schools only*	9	M, L, S, FL	A, V	FL, S, SS, R
Estonia	General only	Yes, for public and government-dependent private schools only	9	M, L	S, SS, FL	
France	All programmes	Yes, for all schools	9	M, L, S, SS, FL, T, A, R		
Germany	General only	Yes, for all schools	9, 10	M, L, FL		
Ireland	All programmes	Yes, for public and government-dependent private schools only	11	M, L, S, SS	A, R, V, O	
Italy	All programmes	Yes, for all schools	8	M, L, S, FL, A		
Netherlands	All programmes	Yes, for all schools	12	L, FL, V	M, S, SS	
Norway	All programmes	Yes, for all schools	10			M, L, FL
Poland	General only	Yes, for all schools	9	M, L, S, SS, FL		
Portugal	General only	Yes, for all schools	9	M, L		
United Kingdom (Scotland)	All programmes	No (but 100% administer them)	11		M, L, S, SS, FL, T, A, R, V	
United States	All programmes	Yes, for public schools only (mandated at the state level)	6, 7, 8, 9	M, L, S		

Notes: Subjects assessed: M: mathematics, L: national language or language of instruction, S: science, SS: social studies, FL: modern foreign language, T: technology, A: arts, R: religion, V: practical and vocational skills, O: other.

All students: all students take the test; Students choose: students can choose to take the test in this subject; Sample: sample or selection of students take the test in this subject.

*) Denmark: 95% of government-dependent private schools also administer the examination.

Source: OECD (2012a), *Education at a Glance 2012*, adjusted and complemented with information supplied by countries participating in the OECD Review. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Assessment criteria and scoring rubrics

In order to help markers make a judgement about the extent to which standards have been met, many education systems have developed additional documents that detail and illustrate different levels of proficiency. Such guidelines can help create a shared understanding regarding what constitutes adequate, good and excellent performance in

different subject areas. They can also provide information regarding the level of performance required for a particular mark. A range of tools can contribute to facilitate the assessment of student work against national curricula, standards or learning progressions:

- **Performance criteria** refer to guidelines, rules or principles by which student performances or products are judged. They describe what to look for in student performances or products to judge quality.
- **Rubrics** refer to scoring tools containing performance criteria and a performance scale with all score points described and defined.
- **Exemplars** refer to examples of actual products or performances to illustrate the various score points on a scale.

Criteria or rubrics may be part of the national curriculum or syllabi, or be provided in a separate document. In Finland, for example, descriptions of good performance and criteria for final assessment are determined within the core curriculum. While the marking scale is from 4 (weak) to 10 (excellent), the national core curriculum only defines criteria for the mark 8 (good) for each subject at transition points. These criteria are intended to help teachers focus assessment on core knowledge and competencies and understand the meaning of "good learning" at different stages. In Ireland, assessment criteria have been developed in secondary education for each subject that is taken in national examinations (the Junior Certificate and the Leaving Certificate). These criteria are closely linked with the syllabus for each subject and are usually an integral part of the discussion during the development of a new syllabus or course. In the Netherlands, benchmarks in literacy and arithmetic were introduced in 2010 to set a common framework of expectations and learning trajectories across all educational tracks in these two core subjects. Each benchmark provides a general description, a description of the tasks that students should be able to perform and the criteria these tasks have to meet, for two levels: a fundamental level that all students should meet and a more advanced level for gifted students (Scheerens et al., 2012).

Information from the education systems participating in the OECD Review indicates that the development of rubrics that detail assessment criteria for systematic and widespread use are not yet common place across countries. In Denmark, for example, the Common Objectives provide goals that articulate the knowledge and skills that "teaching should lead towards" but appear to lack performance standards that describe concrete learning outcomes that students should achieve and criteria for assessing these (Shewbridge et al., 2011). In Norway, school owners have expressed concern about the lack of standards concerning the competencies required for a particular mark and the potentially resulting unfairness in teacher marking (Norwegian Directorate for Education and Training, 2011, in Nusche et al., 2011a). In Sweden, new curricula were introduced at all ISCED levels in 2011 to provide clearer goals and guidance for teachers regarding expected learning outcomes for different subjects and year levels. This was in response to concerns that the learning goals provided by the previous core curriculum remained too vague and led to inequities in teacher marking (Nusche et al., 2011b). In the Slovak Republic, while national performance standards were recently introduced, the level of detail of the criteria to measure achievement of these standards varies among subjects (Hajdúková et al., forthcoming). In the French Community of Belgium, the Inspectorate noted large differences in the required performance levels across schools, despite the existence of common competency references (Blondin and Giot, 2011).

In many settings, assessment criteria are defined at the teacher or school level. In the Czech Republic, for example, criteria are defined by each school and reflected in the respective School Education Programme. As a result, even though all schools may use the same marking scale, they have different marking criteria (Santiago et al., 2012b). As in many other countries, it is not necessarily common practice for teachers to specify assessment criteria in full detail and to inform students of them in advance. In Portugal, there is a strong national focus on developing assessment criteria. At the same time, while national guidelines exist for the development of criteria, schools have some flexibility, for example in the weightings they can assign to different components. Within the nationally defined guidelines, schools are autonomous to specify and publish their own assessment criteria. These criteria are shared with students and guardians and are used to support internal summative assessment as well as decisions about student progress or additional support required (Santiago et al., 2012a).

Qualification frameworks

At the secondary level, several education systems have introduced qualifications frameworks to provide clarity to stakeholders within and outside educational institutions regarding the academic and vocational qualifications students can obtain and the associated competencies they need to acquire. A qualifications framework can be defined as a rank order of qualification levels, allowing each qualification to be assigned to a specific rank. It classifies qualifications according to a set of criteria for levels of learning achieved (OECD, 2007, 2010b). In Europe, the development of the European Qualifications Framework has encouraged the development of national frameworks, in line with the European framework. Education systems such as the Flemish Community of Belgium, Hungary, Ireland, Spain and the United Kingdom have recently introduced such frameworks (OECD, 2010b).

Qualifications frameworks vary across education systems in the level of prescription they provide. In some countries, including the United Kingdom, New Zealand and South Africa, they have a strong regulatory function with common rules across all qualifications. In other education systems, such as Australia and Scotland in the United Kingdom, they are more loosely defined. In these cases, the framework provides more of a map of qualifications available and allows room for differences at the local level (OECD, 2010b). While the implementation of qualifications frameworks is often challenging, there are important benefits in terms of providing clear references regarding the level of competencies that should be associated with different qualifications and how the different qualifications relate to each other. Such transparency in progression pathways at the secondary level can facilitate students' progression in education, while at the same time acting as a quality assurance mechanism for qualifications offered by different providers (OECD, 2010b).

Assessment instruments and approaches

Much has been written about the advantages and disadvantages of different assessment instruments and the type of learning that different instruments can capture. Looney (2011b) describes four broad groups of assessment approaches frequently used in standardised assessments:

- *Multiple-choice assessments* present students with a set of alternative answers. Students must choose one answer from this set. They are machine-scored and provide comparable data on student results. Well-designed multiple-choice items

can assess higher-order knowledge, but they cannot assess broader skills and competencies such as the ability to develop an argument or communicate effectively. Poorly designed multiple-choice tests may be prone to measurement error, for example when students misinterpret questions or make random guesses.

- *Adaptive assessments* are computer based and, as implied by their name, adapt questions to the test-taker. Students who answer questions correctly are directed to a more difficult question and those answering incorrectly receive an easier question. Such adaptive tests can provide more fine-grained information on student performance than traditional multiple-choice tests. However, since not all students respond to the same questions, it is not possible to compare student performance. Also, adaptive tests require a very high number of test questions, which contributes to higher development costs.
- *Performance assessments* are a range of alternative assessment approaches developed partly as a reaction to the widespread use of multiple-choice tests in some countries. In other countries, there is a long tradition of performance assessments and only very limited use of multiple-choice tests. The main characteristic of performance assessment is the intention to assess a range of integrated knowledge, skills and attitudes by asking students to perform a task rather than to provide a correct answer. They may include tasks such as essays, oral presentations, portfolios, experiments and group work. Such assessments tend to capture a broader range of curriculum goals and may be more effective at measuring complex competencies. However, there are concerns regarding the reliability of these assessments. Since scores are typically awarded by human raters, there may be variability in their judgements. Such assessments may also be more expensive to administer and score.
- *Computer-based performance assessments* can potentially assess complex performances by using information and communication technologies (ICT), such as simulations, interactivity and constructed response formats. In some countries, especially in the United States, research has been conducted to develop increasingly sophisticated ICT programmes that are able to score open-ended performances. Such assessments are not yet widely used and still in the early stages of development. They may help address concerns related to the reliability of human raters on the one hand and related to the validity of multiple-choice assessments on the other.

Instruments used in standardised assessments and examinations

OECD education systems use a range of different formats and approaches for standardised central assessments of student learning. Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b (Annexes 4.A1 and 4.A2) provide information on the assessment formats applied in standardised assessment in mathematics and the language of instruction. As is typical for standardised assessment, the vast majority of education systems use written formats for their central assessments and examinations. The instruments used do not differ very much between assessments in mathematics and assessments in the language of instruction within education systems. This section provides information on assessment formats used at the lower secondary level (ISCED 2).

In standardised central examinations that have formal consequences for individual students, the most frequently used assessment formats are open-ended written tasks. Multiple-choice items are also frequently used, especially in examinations in the language

of instruction. A few education systems also use closed-ended short-answer formats. Denmark uses more performance-based formats, namely oral presentations, oral questions and answers and project presentations as part of the standardised national examinations (in the language of instruction only). In Italy, the school-based part of the examination also includes oral questions and answers and oral presentations. Estonia uses performance-based tasks as part of the mathematics examination. In the case of central examinations, the strong focus on written tasks appears to be related to concerns about reliability and fairness in assessment and marking of standardised assessments. In many countries, these centrally designed standardised components are complemented by non-standardised parts of the examinations that are locally designed and marked. While there are limits to what any centrally administered standardised assessment can assess, it is often expected that the assessment of a broader range of skills and competencies happens in such local assessments.

In central assessments that do not have formal consequences for students, multiple-choice tests are by far the most frequently used assessment format in both mathematics and the language of instruction. Closed-format short-answer questions and open-ended writing tasks/open calculations are also frequently applied. Many countries use a mix of these three formats. Only a few countries including Austria (language of instruction only), Iceland, New Zealand and Sweden use oral question and answer formats. Only Austria (language of instruction only), Finland (language of instruction only) and Sweden require students to perform oral presentations as part of the national assessments. The use of performance tasks is also limited; such tasks are applied in central assessments in the Flemish Community of Belgium, the French Community of Belgium (general programmes only), and New Zealand. The limited use of oral and performance-based tasks in central assessments might be explained by the fact that such assessments serve primarily for monitoring and comparing student results across regions and schools and the purpose is to obtain highly reliable and easily comparable scores. The administration of tasks that require one-to-one assessment situations would also be more costly and time-intensive than written tests.

Moving towards more innovative forms of standardised assessment

While central assessment systems tend to rely predominantly on traditional paper-based assessment formats, there are several interesting examples of assessments that attempt to capture a wider range of competencies. Innovative performance-based assessment formats offer significant potential to signal the learning goals that are valued in the education system. They can also model to teachers the next generation of assessment formats that can measure and support broader learning.

To help teachers use innovative forms of assessment, some countries have developed specific assessment instruments that are made available for teachers to use when they see fit. Several countries have developed on-demand assessments, where teachers can draw from a central bank of assessment tasks and ask students to take the assessment when they consider that they are ready. This gives teachers control of the timing of assessment. In Australia, for example, the government's Online Diagnostic Tools initiative provides teachers with access to online resources that can help assess student progress and provide links to relevant digital learning resources to help them improve. Assessment instruments may also be developed by other actors in the education system. In the Flemish Community of Belgium, for example, several specific evaluation instruments have been developed by various institutions to help teachers assess the non-cognitive performances of their students (Box 4.7).

Box 4.7 Assessment formats that measure broader competencies

In the **Flemish Community of Belgium**, a range of institutions including educational centres, academic institutes and umbrella organisations, have developed tools that teachers can draw on to assess non-cognitive aspects of learning, such as attitudes, well-being and involvement. Some of these tools are used quite widely while others are applied rather sporadically and in a limited number of schools or education forms. The most commonly used tool for assessment of non-cognitive performance is the SAM-scale (Scale for Attitude Measurement; *Schaal voor AttitudeMeting*)[7]. The SAM-scale measures students' attitudes and has been developed to assess, guide, stimulate and orientate students. By means of this tool a teacher can determine to what extent a pupil scores high or low for certain attitudes, e.g. flexibility, diligence and responsibility.

In **Finland**, "learning to learn" skills are considered to be central to each student's development. These are actively promoted as core elements in achieving lifelong learning and include a student's capacity for independent and self-motivated learning, problem-solving and the ability to evaluate his/her own learning and related strategies. There is a clear pedagogical goal in all compulsory education subjects for students to develop "learning to learn" skills. To evaluate and promote the importance of such skills, national sample assessments were developed by the Centre of Educational Assessment at the University of Helsinki to evaluate "learning to learn" skills in Years 3, 6 and 9 of compulsory education.

In **New Zealand** primary schools, progress towards the achievement of national curriculum goals is measured via the National Education Monitoring Project (NEMP). NEMP is designed to be as well aligned as possible with the curriculum by incorporating competency and value elements. Many of the NEMP assessment tasks are performance-based, requiring students to transfer learning to authentic close-to-real life situations. There are different assessment situations including one-to-one interviews, work stations and teamwork. As the assessment does not carry high stakes for students it is particularly important that tasks are meaningful and enjoyable to them. The assessment provides rich information on the processes used by students to solve problems or conduct experiments. Most assessment tasks are carried out orally so as to analyse what students can do without the interference of reading and writing skills. Some of the tasks are videotaped to allow for an in-depth analysis of student responses and interaction with teachers. NEMP also assesses students' cross-curricular skills, and attitudes towards the learning areas being assessed. Students' enjoyment of particular assessment tasks is also surveyed. For instance, 82.5% are reported as enjoying hands-on tasks, versus around 50% for paper and pencil tests. While NEMP is designed for system monitoring, examples of previous assessment tasks are available for teachers and may be used in the classroom. This can help teachers estimate how their own group of students compares to national assessment results.

Sources: Flemish Ministry of Education and Training (2010); Finnish Ministry of Education and Culture (forthcoming); Nusche et al. (2012)

Other countries have developed sample-based surveys that cover larger parts of the curriculum including non-cognitive elements (see Box 4.7 for examples from Finland and New Zealand). While these primarily aim to provide information at the education system level, they also give important insights to participating students and schools regarding valued key competencies. In addition, where the assessment tasks of previous years are published, these can serve as example tasks for formative assessment in all schools. In New Zealand, for example, such use of previous assessment tasks is encouraged by the education authorities.

As will be discussed further below, developments in information and communication technology (ICT) also offer new avenues for assessing broader competencies on a larger scale. Of course, ICT-based technology is not a prerequisite for large-scale assessment of students' open-ended performance, but they may make such assessments considerably more affordable.

Instruments used in teacher-based assessment

In the education systems participating in the OECD Review, teachers are largely autonomous in the choice of internal assessment instruments within the limits of the local or national curriculum. Typically, there are no central regulations regarding the particular assessment instruments to be used in measuring the achievement of learning goals. Given the diversity of approaches applied by teachers in different subjects, year levels, regions, schools and classrooms, it is difficult to characterise and compare "typical" assessment formats used within a system. However, the information collected through Country Background Reports and OECD Country Reviews can give some initial indications and observations regarding internal assessment trends within countries.

It appears that, traditionally, teachers in all education systems have relied as a minimum on written assessments and examinations and certain oral assessments when measuring student performance in core subjects. Written assessments may take many different forms, such as multiple-choice tests, quizzes, dictations, exercises, short and long constructed response tasks and essay questions, while oral examinations typically take the form of question and answer exercises. The traditional reliance on rather formal assessment formats may be explained by the way teachers' assessment role has long been conceptualised in many countries. In many settings, the understanding of assessment has been one of summative intention, serving primarily to provide reliable and comparable results and rankings of students in order to make decisions about placement, advancement and selection. In many countries, the checking of homework assignments and observation of classroom participation may also be part of classroom-based assessment. Depending on the subject, particularly for subjects requiring more practical evidence of student performance such as physical education, music, arts and sciences, it is reported that teachers have traditionally relied also on broader formats such as practical demonstrations, performances and experiments.

Teachers across education systems participating in the OECD Review continue to draw on the abovementioned assessment approaches. At the same time, in many settings, teachers are going beyond traditional assessments and report about their use of a much broader mix of assessment approaches including more sophisticated assessment types aiming to capture a broader range of integrated knowledge and skills. In most countries where OECD Country Reviews were organised, teachers reported using various forms of alternative assessment, including products (e.g. written essays; laboratory reports), performance (role plays, experiments, presentations) and portfolios. Such a range of assessments using *product, portfolio and performance*, are also known as "3P assessments" and reflect an emphasis on assessing "higher-order" thinking skills in authentic ways (Madaus and O'Dwyer, 1999; Stiggins, 1987). In several settings, teachers also reported relying more on observing and recording student achievement as it occurred and mapping progress through the collection of student work samples over time.

Promoting the use of innovative assessments by teachers

To support a balanced approach to teacher-based assessment, several education systems have recently launched central initiatives to help establish a new culture of assessment focussed on more performance-based approaches to measuring student learning. In France, for example, personalised competency monitoring systems were introduced for teachers to record their students' progress in acquiring core knowledge and competencies (Box 4.8).

Box 4.8 Continuous teacher-based assessment of student competencies

In **France**, the 2005 orientation law highlights the requirement for schools to ensure that all students achieve a common core (*socle commun*) of knowledge and competencies necessary for their success in school as well as their personal and professional future. In this context, teachers are required to assess their students' progress towards the common core of competencies using the students' Personal Competency Booklet (*Livret Personnel de Compétences, LPC*) throughout their compulsory schooling. A range of items that students should have acquired at key stages of education are defined through the legislation, and teachers attest students' mastery of these items on the Personal Competency Booklet. Items are validated by classroom teachers in primary education and by the class councils in lower secondary education, with the school leader approving the overall mastery of the common core of competencies at the end of lower secondary education. While it is not obligatory for students to validate each item (this is left to the discretion of teachers), it is mandatory for them to validate each of the seven key competencies defined in the 2005 orientation law. To support this assessment process, all teachers have received guidance material defining the different competencies students are expected to achieve and providing a number of assessment tools and reference criteria to undertake the assessment.

Source: Dos Santos and Rakocevic (2012).

In other systems, support for broader assessment formats takes the form of central frameworks or guidelines for assessment suggesting a range of different assessment approaches to teachers so as to ensure that different types of learning are given adequate attention across the curriculum.

In Austria, for example, a new secondary school type (the "New Secondary School") was launched in 2008 as an inclusive school to avoid the early tracking of students into different school tracks after Year 4. This attempt at a structural reform is accompanied by pedagogical reforms such as new ways of individualisation and differentiation of instruction, social learning and integration of students with special needs. In line with the focus on promoting innovative pedagogical approaches, the new secondary schools are required to use more formative forms of assessment and to focus on the assessment of performances, through, for example, self-observation, self-assessments and portfolios (Specht and Sobanski, 2012).

In the Flemish Community of Belgium, the central education authorities are promoting a shift towards a "broad assessment culture" which includes a focus on formative assessment and new assessment approaches. It implies the use of "alternative" (compared to tests) assessment approaches including observation, portfolios, reflection sheets and self- and peer-assessment. It is also being highlighted that it is more important to report on student progress compared to previous achievements rather than on absolute performance. While it takes time to bring about such a culture shift systematically across all schools, progress has been made in promoting alternative forms of assessment, in particular in primary education, vocational education and special needs education (Flemish Ministry of Education and Training, 2010).

In Ireland, the Primary School Curriculum provides for the use of assessment approaches such as teacher observation, teacher-designed tasks and tests, work samples, portfolios and projects, curriculum profiles, diagnostic tests and standardised tests. The National Council on Curriculum and Assessment (NCCA)'s guidelines on assessment present these and other assessment approaches as a continuum of approaches moving from those that are pupil-led such as self-assessment and conferencing to those that are more teacher-led such as teacher observation, teacher-designed tasks and tests, and standardised testing. Assessment approaches such as portfolio assessment, concept

mapping and questioning appear towards the middle of such a continuum. In secondary schools, informal methods of assessment include classroom tests administered at the end of topics or segments of the syllabus to assess students' knowledge and mastery of particular skills, informal observation by the teacher, evaluation of homework, and informal analysis of students' language and social development (Irish Department of Education and Skills, 2012).

In Mexico, the national curriculum (Study Plan) states that rubrics, checklists, registries of observations, written pieces of work, team projects, conceptual maps, portfolios and written and oral tests should be used. It also requires that students should be frequently involved in self-assessment and peer assessment activities (Santiago et al., 2012c).

In New Zealand, also, the importance of using multiple sources of evidence for effective assessment is emphasised at both primary and secondary levels. Schools are required to use a range of assessment practices to measure students' progress and achievement in relation to the national curriculum and standards. Thereby, they are encouraged to use a diversity of approaches in school-based assessment. The focus on broad assessment is further emphasised by the reliance of National Standards on overall teacher judgements (OTJ) rather than a national standardised test (Nusche et al., 2012).

In Slovenia, the principles for assessment and examinations are specified in Rules on Examination and Assessment of Knowledge for the different levels of education. Among other things these principles require teachers to use a variety of forms and methods of verification and assessment of knowledge; to take into account the ability to analyse and interpret creativity and the ability to use knowledge; to allow students a critical reflection and insight into the acquired knowledge and to contribute to the democratisation of relations between students and teachers. In primary education, student achievement is assessed continuously in written, oral, artistic, technical and practical forms, as well as through tests set by teachers. In upper secondary education, teachers assess students' oral answers, written tests, visual arts and other works, seminars, projects and presentations (Brejc et al., 2011).

In Singapore, the "thinking schools, learning nation" initiative was introduced in 1997. The explicit focus of the related reforms was on developing a creative and critical thinking culture within schools, developing an inquiry orientation among teachers and students and assessing students on these competencies. These initiatives were also linked to a commitment to integrating technology into all aspects of education. New subjects, such as "project work" and "knowledge and inquiry" were introduced, along with requirements to use specific performance-based assessments in these subjects. Many courses of the reformed curriculum include applied examination elements that allow students to show their problem-solving and thinking skills (Darling-Hammond, 2010) (Box 4.9).

In some countries, assessment approaches developed in the vocational education and training (VET) sector or in second-chance education seemed to be ahead of the general sector in terms of paying attention to wider competencies and making efforts to assess these through sophisticated assessment approaches.

In Finland, for example, all vocational qualifications include skills demonstrations, which form part of student assessment. The demonstrations take place in a practical and authentic work situation, where the student shows how well he or she has attained the objectives and aims set for vocational competence in the Core Curriculum. In 2004, the Finnish National Board of Education commissioned a review of the effect of skills

demonstrations on the quality of vocational education and training. The study concluded that skills demonstrations have had several positive effects. The system was found to help assure the level of students' learning, respond to the needs of the labour market and enhance the quality of VET education, because the feedback received from professionals could be used in the further development of training and teaching. Overall, the review indicated that the skills demonstrations had a positive effect on students' motivation and aptitude to learn and increased the value of VET in the labour market (Finnish Ministry of Education and Culture, forthcoming).

Box 4.9 Singapore: Creating "thinking schools"

In Singapore, recent reforms have changed the curriculum and assessment system to make it more explicitly focused on creativity and independent problem solving. Curriculum and assessment guidelines encourage teachers to use a variety of assessment formats. The Ministry has developed support tools for teachers such as:

- The Strategies for Active and Independent Learning (SAIL) which aim to support learner-centred project work and provide assessment rubrics to clarify learning expectations. All schools have received training for using these tools.
- The Ministry's 2004 Assessment Guides for primary and lower secondary mathematics which contain resources, tools and ideas to help teachers incorporate strategies such as mathematical investigations, journal writing, classroom observation, self-assessment and portfolio assessment into the classroom. The Institute of Education has held a variety of workshops to support learning about the new assessments and integrated the new strategies into teacher development programs.

In addition, Project Work was introduced as an interdisciplinary subject that requires students to draw knowledge and apply skills from across different subject domains. The requirements for project tasks are centrally set by the Singapore Examinations and Assessment Board. The tasks are defined in a broad way so as to allow students to carry out a project of their interest while meeting the following task requirements: *(i)* it must foster collaborative learning through group work, *(ii)* every student must make an oral presentation; and *(iii)* both product and process are assessed and there are three components to assess them: a written report, an oral presentation and a group project file to which each group member submits three documents related to snapshots of the processes involved in the project. About 12 000 students complete this task annually. Assessment is school-based and criterion-referenced. While task setting, conditions, assessment criteria, achievement standards and marking processes are externally specified by the Singapore Examinations and Assessment Board (SEAB), the assessment of all three components of Project Work is carried out by classroom teachers, using a set of assessment criteria provided by the board. All schools are given exemplar material that illustrates the expected marking standards. The Board provides training for assessors and internal moderators. Like all other assessments, the marking is both internally and externally moderated.

In a recent paper, Koh et al. (2011) analyse the Singaporean experience and highlight the need to bring about fundamental changes of school cultures and teacher and student dispositions in order to make these assessment reforms work in practice. They find that despite policy change teachers often remained reliant on conventional assessment and were sometimes ill-prepared to implement authentic assessment. In a two-year empirical study on teacher professional development in authentic assessment and use of rubrics, they found that teachers were able to improve the quality of classroom assessment after participating in certain types of professional development. Not surprisingly, active and collective participation in ongoing, sustained professional development was more effective than ad hoc or 1-2 day workshops in building teachers authentic assessment capacity.

Sources: Darling-Hammond (2010); Koh et al. (2011).

In Portugal, with the so-called New Opportunities programme, there have been a number of initiatives to extend the educational provision in schools to students who may have left school, or may be at risk of leaving school, and to adults who might not have completed compulsory education. These new programmes have been accompanied by the development of approaches to assessment focused on motivating students, giving high-quality feedback, and including the active participation of learners in the assessment process. A key feature of these arrangements and approaches is their location close to the learning process and to the learner. Thus, assessment tends to occur immediately after the completion of a module or portion of a course rather than at the end of a year or cycle. The use of approaches beyond written tests, such as a performance assessment, puts the learner and learning at the centre of the assessment process. The need to use assessment to motivate learners to learn, rather than to simply engage with the assessment or test, is acknowledged as fundamental by those working in this sector (Santiago et al., 2012a).

While innovative assessment of practical skills is essential in vocational education and training, OECD (2010b) argues that such assessment should be embedded within a standardised central assessment framework. The advantage of having a standardised framework for assessment in the VET sector is to ensure that all those with a qualification have the same mix of competencies, in particular where there is substantial variation among individual VET institutions and companies offering apprenticeships. There are a range of potential approaches to ensuring common standards in assessment in VET, such as periodic inspections of VET institutions, inspections of examination bodies, random evaluation of student performance, self-evaluation of providers, peer reviews, clear central guidelines for locally developed examinations and availability of a range of standardised assessment tools (OECD, 2010b).

Ensuring consistency of marking through moderation

A key strategy to increase the reliability of assessment and marking across different settings is to systematically implement moderation procedures. Moderation refers to a set of approaches that aim to ensure the quality and comparability of assessment judgement. It may involve teachers cross-marking each other's assessments or discussing student performance in groups, or a competent external organisation systematically checking school-based marking.

While in many settings, moderation occurs informally within and between schools and may not be documented, some education systems have introduced systematic arrangements for moderation. This is particularly the case in education systems where centrally developed examinations with high stakes for students are corrected and marked locally by teachers. In the French Community of Belgium, schools are in charge of marking their own students' examinations and they decide autonomously whether students' examinations are corrected by their own teacher, another teacher or a group of teachers. Guidance materials are available to support schools in this task. In France, while teachers examine their own students through continuous classroom assessment, teachers from another school are responsible for marking written examinations leading to diplomas or certification. In Denmark, centrally appointed external examiners correct examination papers and are assisted through national guidance materials such as performance criteria, exemplars, rubrics and keys. There is also moderation of marking by external examiners who attend oral examinations. In the Netherlands, examinations are corrected by the students' own teacher and moderated by a teacher from another school using a central scoring protocol. The school boards are responsible for the proper

handling of the procedures. In case of disagreement, external moderation by a competent body is provided (Tables 4.A1.4, 4.A2.4a and 4.A2.4b, Annexes 4.A1 and 4.A2).

In Australia, while each state and territory has its own system and own set of procedures for developing and approving courses in secondary education, most combine student performance on external exams at the end of Year 12 with moderated, teacher judgements of coursework performance to arrive at scores for senior secondary certificates and high school completion. In Queensland, the examination system is school-determined and based, but achievement standards and scoring are externally moderated. Moderation processes for the Senior Certificate (Year 12) involve subject-based panels of expert teachers providing advice to schools on the quality of their assessment programme and their judgements of quality of student performance based on sample portfolios. The system involves follow-up with schools where panels identify issues regarding assessment and standards. There is negotiation of the final results to be recorded on the Senior Certificate (Sebba and Maxwell, 2005 in Santiago et al., 2011). Similarly, procedures adopted by educational jurisdictions and particular schools for moderating internal summative teacher judgements (so-called A-E ratings) also facilitate common understanding of year level proficiency standards and foster the development of professional learning communities that can provide crucial support for improving opportunities for student learning and building teacher capacity (Santiago et al., 2011).

In New Zealand, an external moderation system is also in place to ensure the dependability of internal assessments in Years 11-13. The New Zealand Qualifications Authority directly checks the quality of internal assessment through a sampling approach. Schools are required to submit 10% of internally assessed student work for NZQA moderation to make sure the assessment is appropriately aligned with standards. The moderation process does not affect the marks assigned to assessment samples by teachers, but is intended to provide feedback to teachers and to inform future assessment policy development at the system level (Nusche et al., 2012).

Use of technology in assessment

In recent years, the potential of information and communication technologies (ICT) to influence and shape assessment approaches has been increasingly recognised across OECD countries. Binkley et al. (2010) describe two key strategies regarding the use of ICT in assessment. First, the "migratory" strategy refers to the use of ICT to deliver traditional assessment formats more effectively and efficiently. Second, the "transformative" strategy refers to the use of ICT to change the way competencies are assessed and develop formats that facilitate the assessment of competencies that have been difficult to capture with traditional assessment formats (Binkley et al., 2010; Ripley, 2009; European Commission, 2011a).

Information collected from education systems participating in the OECD Review shows that the use of ICT for assessment has not yet become common practice internationally. In the few systems where technology is used for standardised central assessments or examinations, it is mostly done in a "migratory" perspective. Some countries, including Australia, the Czech Republic and New Zealand use technology in national examinations for data management purposes, such as data sheet scanning, inputting marks or results management, while examinations remain paper-based. In a few countries, such as Slovenia, computer-based technology is used for students with special educational needs. Some education systems, namely the Flemish Community of Belgium, the Czech Republic, Luxembourg (ISCED 2), New Zealand, Norway and Northern

Ireland in the United Kingdom use computer-based uniform technology for the actual administration of central assessments (and for central examinations in the case of Norway).

Denmark is the only country participating in the OECD Review which reported using computer-based adaptive technology for its national assessments in Danish (reading), English, mathematics, biology, geography and physics/chemistry for different year levels. In these adaptive assessments, test items are selected sequentially according to a student's performance on the previous test items. This makes testing more efficient as more fine-grained information can be obtained in less testing time. As reported in the Dutch Country Background Report, the Cito pupil monitoring system in the Netherlands also contains computer-based tests some of which are adaptive (Scheerens et al., 2012).

In Australia, through the governments' Online Diagnostic Tools initiative, the potential to deliver national assessment programmes on line is also being explored through detailed research and consultations on the measurement and assessment effects of online testing and through discussions with school authorities regarding technical and delivery considerations. The delivery of national online assessment offers the opportunity to further integrate ICT into teaching and learning and provide more individualised diagnostic assessment of student progress.

Research and development in ICT-based assessment

While the systematic use of ICT to transform central assessment systems is still limited, many public and private actors are increasingly investing in research and development in this area. Several innovations in assessment around the world are taking advantage of recent advances in ICT. Increasingly sophisticated ICT programmes that score "open-ended performances", such as essays, are under development (see Chung and Baker, 2003; Chung et al., 2001; Herl et al., 1999, in Looney, 2009). These programmes use natural-language processing, artificial intelligence and/or information retrieval technologies to detect textual features of essays (for example, variety in use of syntax, quality of content and organisation of ideas). These ICT models are still in the relatively early stages of development however, and while they may facilitate scoring of large-scale assessments, cannot replace human raters. Further studies are also needed to determine the validity and reliability of different automated essay scoring tools (Wang and Brown, 2007, in Looney, 2009).

Technology-based assessments may also incorporate simulation, interactivity and constructed response formats. For example, students may use the multimedia functions of ICT to show how they would perform a science experiment or other problem solving tasks (e.g. the Virtual Performance Assessment project at Harvard University, Box 4.10). They may use Internet–based programmes to "predict-observe-explain" specific concepts. Or, they may develop concept maps using online tools to show their understanding of processes. The student's map can then be scored by comparing it against an expert map (Bennett, 2001, in Looney, 2009). There are also examples of web-based peer assessment strategies (Binkley et al., 2010). Research is also undertaken in the area of assessments that can measure collaborative problem-solving. For example, students can send information and documents to each other and work on tasks together using ICT (Binkley et al., 2010).

In another example of ICT-based assessment, students are scored on their use and judgement of information on the Internet. Students may obtain feedback on their work in real time. Examples of tests that track student activities on a computer while answering a

question or performing a task are the iSkills test developed by the Educational Testing Service (ETS) in the United States, and the ICT literacy test which is part of Australia's National Assessment Program (Box 4.10). Test developers have found that students respond positively to Internet tasks (searching, and judging the quality of on-line information) and are engaged, even in difficult, open-ended tasks (Lewis, 1998, in Looney, 2009).

Box 4.10 Sophisticated ICT-based assessments

The **Virtual Performance Assessment (VPA)** project at Harvard University (*http://vpa.gse.harvard.edu/*) in the United States uses innovations in technology and assessment to measure students' ability to perform scientific inquiry to solve a problem. The virtual assessments use similar design as videogames to support students' experimentation and problem-solving skills. Participants take on the identity of a virtual scientist and can walk around the environment, make observations, gather data, and solve a scientific problem within a context. The student's arguments are expressed through various digital media including concept maps, data collection tools and conclusion tools for submitting hypotheses and causal statements. The student's actions are stored in logfiles and can be used to track students' inquiry trajectories.

The **iSkills test** developed by the Educational Testing Service (*www.ets.org*) in the United States intends to measure constructs not directly accessible through conventional means, testing the ability to use technology as a tool for cognitive purposes. The iSkills test intends to measure students' critical thinking and problem-solving skills in a digital environment. In a one-hour exam real-time, scenario-based tasks are presented that measure an individual's ability to navigate, critically evaluate and understand the wealth of information available through digital technology. The programme provides individual and group data for use in student evaluation and placement.

The **National Assessment Program** in Australia includes an assessment of students' ICT literacy. It is designed as an authentic performance assessment. The assessment is intended to mirror students' typical "real world" use of ICT. In the 2005 and 2008 rounds of the assessment, students completed tasks on computers using software that included a seamless combination of simulated and live applications. Some tasks were automatically scored and others (those that resulted in information products) were stored and marked by human assessors. The tasks (items) were grouped in thematically linked modules, each of which followed a narrative sequence covering a range of school-based and out-of-school based themes. Each module typically involved students collecting and appraising information as well as synthesising and reframing the information. The assessment involved a number of modules so as to ensure that the assessment instrument assessed what was common to the ICT literacy construct across a sufficient breadth of contexts (MCEECTYA, 2008). The 2011 ICT Literacy report is now also available.

The New Zealand Ministry of Education (2011) has developed digital portfolio guidelines "for beginners". These guidelines aim to provide information on e-portfolios to non-technical users and can help school leaders consider the place of e-portfolios in their school's educational strategy. The guidelines also provide an overview of available tools and case studies of schools having implemented the use of e-portfolios. The Ministry of Education has also supported the development of the e-portfolio service **My Portfolio (Mahara)** (*http://myportfolio.school.nz/*), which provides a personal learning environment to record and showcase evidence of achievement, manage development plans, set goals, and create online learning communities. New Zealand Schools can register free of charge. Most New Zealand universities have introduced MyPortfolio to their teacher training programmes and some are now familiarising their schools' advisers with MyPortfolio. The Ministry of Education is also offering "taster" sessions to groups of teachers with an interest in using MyPortfolio.

Sources: Binkley et al. (2010); European Commission (2011a); MCEECTYA (2008), New Zealand Ministry of Education (2011).

Recent developments in ICT are relevant not only for standardised assessments, but they can also influence regular assessment practices in the classroom. For example, there has been increased interest in using digital portfolios (or e-portfolios) across countries (McFarlane, 2003; Binkley et al., 2010; Pepper, 2011). While portfolios have been used in many countries for some time, the use of digital tools allows collecting information on student progress in a broader range of formats including text with hyperlinks, video, audio and simulations. Digital portfolios also make it easier for teachers to comment on assignments and track student progress. Students' own work with the digital portfolios can enhance their skills in learning to learn, ICT literacy and self-monitoring (Binkley et al., 2010; Pepper, 2011). The New Zealand Ministry for Education, for example, developed e-portfolio guidelines for non-technical users and supported the development of an e-portfolio service called MyPortfolio (Mahara).

ICT-based assessment also provides opportunities for more equitable and adapted assessment of diverse students. The degree of personalisation possible in ICT-based assessment is greater than in traditional assessment given the range of presentations, response formats and contexts available, for example through item banks, test practices and e-portfolios. The Computers for Pupils initiative developed by the UK Department for Children Schools and Families (now Department for Education), constitutes one attempt to increase equity through e-Assessment. Through providing e-learning and e-assessment in the homes of disadvantaged children, this initiative has increased their participation in assessment (e.g. through handheld and haptic technologies). Furthermore, this initiative has provided students with a platform for social networking and self-help groups. In such communities, students may get in contact with peers that experience the same challenges and provide each other with mutual support (Nayral de Puybusque, forthcoming).

Equity and fairness in assessment design

It is important that assessments allow all students to show what they know and can do without being unfairly hampered by individual characteristics that are irrelevant to what is being assessed (Binkley et al., 2010). Assessment needs to be appropriate for students at the range of developmental levels likely to be in assessed population and sensitive to the needs of particular groups such as cultural minorities, students whose mother tongue is not the language of instruction (L2 students) and students with special educational needs. In the process of developing student assessment, notably standardised testing, it is often unclear whether accessibility for specific student groups such as cultural minorities or students with disabilities has received sufficient attention.

The research literature on bias and equity in assessment is extensive, but it often focuses on specific sub-groups, in particular disabled students and L2 students. To ensure fairness in assessment for all students, it is important to develop frameworks for equitable assessment for the wide range of different sub-groups without privileging one group over another. The development of a broad framework for equity in assessment for *all* children requires central guidelines for orientation and coherence across educational settings, but it should at the same time allow for flexibility and adaptability of practices at the local and school level (Nayral de Puybusque, forthcoming).

The design of assessments is open to a number of risks for equity namely if it is biased for particular student groups. Test bias refers to differential validity of a test between specific sub-groups of students (Sattler, 1992). There are several types of potential bias. For example, *irrelevant context* may bias a test against a certain group of students. Unnecessary linguistic complexity is one example for context-irrelevant bias in

assessment, particularly when testing students who do not speak the language of instruction and assessment at home (L2 students). Linguistic complexity slows the student when reading and listening in comprehension and answering, gives room for misinterpretations, and increases the cognitive load (Nayral de Puybusque, forthcoming).

There may also be *bias in content validity* when the choice of a particular set of knowledge and skills is likely to privilege certain groups of students over others (Klenowski, 2009). In this case, after the general ability level of the two groups is held constant, the test is relatively more difficult for one student group than another (Reynolds, 1998, in Whiting and Ford, n.d.). For example, if asked a question about hockey, a student or group who has never played or watched or had discussions about hockey is at a disadvantage. The lack of exposure and experience in relation to particular content places them at a disadvantage (Whiting and Ford, n.d.). There may also be *bias in item selection*, which is related to how one item is included in the test while another is not. While an overall test may not be biased statistically, a few items in it may be. Finally, the choice of method may also lead to bias for certain groups, depending on their familiarity with the general idea of a test, the motivational context in which the test is taken and the frequently implicit assumptions about appropriate behaviour in such a context. For example, students who are more familiar with multiple-choice tests may have developed better strategies to deal with this assessment method than students who have never been confronted with this format before.

Reviewing equity dimensions in test design

While reliability and validity are necessary conditions for any effective assessment system, one cannot assume that these conditions are met or transferable to all different subgroups of the population. Evidence of differential validity is required to determine whether separate test validities are needed for each group (Shultz and Whitney, 2005).

"Equity scanning" is a statistical approach to address testing bias. This method provides the information and means for test developers to identify and eliminate inequitable elements in a test and ensure its quality and the integrity. Differential Test Functioning (DTF), Differential Item Functioning (DIF) and Differential Distractor Functioning (DDF) constitute techniques that evaluate a test when applied to a heterogeneous group of students. All of these three methods aim to assess the technical quality, i.e. the validity of a test. The different methods look at variation of performance across sub-groups on different levels of bias. DTF examines a test as a whole; DIF analyses a test at the item level; and DDF focuses on distractors (i.e. incorrect options in a multiple-choice test). Where two students with the same knowledge, abilities and skills turn out to perform differently at a test in terms of scores, answer, or preferred choice, these techniques may be applied. However, the effects and benefits of these methods have been debated and equity scanning still constitutes mainly an area of academic enquiry rather than a field of policy implementation (Nayral de Puybusque, forthcoming). Statistical methods are often only applied ex-post and, therefore, do not address the problem of inequity in assessment directly. Some authors have also criticised that these measures are more likely to detect test anomalies rather than detecting real bias (Battiste, in Volante, 2008).

"Judgemental reviews" constitute another technique to eliminate culturally biased items in a test. Teachers from diverse backgrounds may be allowed to screen a test for cultural bias within a review process. However, it is important to bear in mind the significant impact of the respective teacher's subjectivity when evaluating this method. To address both construct and method bias, it has also been proposed to use informants

with a thorough knowledge of the respective culture and language of a sub-group (van de Vijver and Tanzer, 2004). In the "committee approach", a group of people with different areas of expertise (such as cultural, linguistic, and psychological) review a test or a test translation. The major strength of the committee approach is the co-operative effort among members with complimentary areas of expertise. In Australia, consultations were conducted regarding the accessibility of the National Assessment Program – Literacy and Numeracy (NAPLAN) for Indigenous students. The NAPLAN test development process involves the consultation of Indigenous experts in states and territories. They provide specific feedback on the suitability of the test items and the appropriateness of the stimulus materials for Indigenous students. Test trials are also carried out using a sample of students, and analysis of the results is undertaken to ensure that all items are culturally appropriate and free of bias (Santiago et al., 2012).

Providing accommodations to specific student groups

The tension between standardised tests that are supposed to be common across all students and the need to be sensitive to local cultural knowledge and individual student conditions is a difficult one. One option to provide fair assessment, while upholding common standards, is to provide accommodations for specific student groups. Accommodations aim to "remove causes of irrelevant variance in each student's test performance, thereby producing a measure of each student's knowledge and skills that is valid" (Thurlow et al., 2008). Such measures are not expected to decrease learning expectations for these students but rather to reduce or even eliminate the effects of a student's disability or disadvantage in relation to a particular assessment (Thompson et al., 2005).

Hopper (2001) described accommodation through four aspects: presentation, response, setting, and scheduling/timing. *Presentation* refers to alternate modes of access to the information adapted to the student's particular condition (e.g. a disability or a limited proficiency in the assessment language). *Response* accommodation allows for alternate ways of completing an assessment, with the possibility of using other materials or devices. *Setting* accommodation allows for different assessment conditions in terms of assessment location and environment. *Timing and scheduling* accommodation allows for taking a test at a different point in time, for a different length of allocated time (i.e. including extra-time), or its organisation (e.g. reallocation of time lengths depending on the task). One specific method of accommodation can address one or also several of these dimensions. The use of ICT-based assessments has great potential to provide a range of accommodations for specific student groups.

For sensory disabilities (i.e. hearing and visual impairments), the most widely used accommodations lie in the field of presentation and response. Examples for the various measures used include Braille or large-print papers, a reader, audio playback, a writer, audio recording of oral responses, headphones, a sign-language interpreter, or additional time. In the case of physical and cognitive disabilities electronic devices are common, but setting accommodations are most widely used. For students with physical disabilities (including reduced mobility), access to different test centres, to appropriate seating for a student to be able to take medication and computer devices constitute typical forms of accommodation. For students with behavioural disabilities, supervised rest breaks, permission to move around, and additional time, especially for dyslexic students, are widely used. For students with dyslexia or learning disabilities, spellchecker software can also be available (Nayral de Puybusque, forthcoming).

It should be noted that evidence on the benefits of using different types of accommodations in assessment is often lacking or not conclusive. Abedi et al. (2004), for instance, raise caution regarding the use of dictionaries in assessment. Depending on the brand, the content and the vocabulary level of the definitions, dictionaries may be more or less useful to a student. L2 students must be able to understand the definitions and be familiar with the use of a dictionary. For assessment accommodations to be most helpful to students, it is important that they are well aligned with instructional accommodations. To this end, some education systems require proof that the demanded accommodations have been provided and used in class and/or in internal examinations (Thurlow et al., 2008).

Assessment accommodations for second language learners

Accommodation with regards to the language of assessment is a crucial, but at the same time very complex concern. For important language minorities, several countries have attempted to provide a range of options for students to be assessed in their first language. This can be done by administrating the assessment orally involving an assessor who is familiar with the student's linguistic background. Another possibility is to translate or develop assessment instruments in the student's first language. There are several options for doing this: One would consist in translating and adapting most existing tests which are publicly available. Another option would be to develop instruments in the minority language. Conceptual equivalence would be accomplished by having bilingual teachers and experts working together to moderate the test construction in each language. Assessment instruments developed in this context would have no metric equivalence and intergroup comparisons would not be possible. The main benefit of this procedure, however, would be to reduce the risk of any cultural or linguistic bias.[8] A third option would be to find an intermediate solution between the two previous ones, by developing anchor points in assessment instruments developed in two different languages.[9]

However, assessment in a student's first language, whether through oral administration or translation, may entail several limitations and requirements when students are not actually taught in that language. Several authors have argued that first language assessments are useful only if students can demonstrate their performance more effectively in their first language, typically because they have received relevant instruction in that language (Abedi et al., 2004). To achieve alignment between the language of instruction and the language of assessment, a test needs to reflect the terminology and language used in content-area instruction (Abedi, 2004).

Reducing linguistic complexity is another frequently used approach to eliminate bias against L2 students in assessment. Low-frequency vocabulary, passive voice verb constructions and the length of written constructions feature among the main challenges for L2 students (Abedi and Gandara, 2006). Accommodation measures to reduce the linguistic complexity of a test include simplification, modification and translation of test elements (e.g. items, directions) as well as the use of dictionaries, and glossaries. Linguistic modifications of content-based items without changes in the content-related terminology represent one possible intervention.

At the same time, it is important to bear in mind that the state of the empirical evidence on the validity, effectiveness and efficiency of many accommodation measures varies and is at times contradictory. In the various states of the United States, Rivera et al. (2006) identified 75 different accommodation tools allowed for L2 students (Rivera et al., 2006). Out of these, only 44 were recognised to fit student's needs. Abedi (2004) judged only 11 measures as effective and appropriate. Another field where progress needs to be made

relates to the diversity of the L2 sub-group and equity within this group. In most countries, it is not cost-effective to develop translated assessments in all the many languages spoken by learners at home, which introduces questions of fairness and equity for smaller language groups. In the United States, for example, while the large presence of the Spanish-speaking minority has brought benefits for L2 students as a whole, equity towards smaller L2 groups is an important issue that should not be neglected (Abedi et al., 2004).

Improving equity through multiple assessment opportunities

Several studies indicate that certain formats of assessment may advantage or disadvantage certain student groups (Gipps and Stobart, 2004). Since it is very difficult to make assessment wholly inclusive and neutral – any format, content, construct, and method may be biased in some direction – a mix of different versions of format, content and construct may help ensure fairer assessment. High-stakes decisions about students should not be based on the results of one test alone. An important approach to offering fairness is to collect multiple data, use a range of assessment tasks involving a variety of contexts, response formats and styles and draw on this comprehensive approach to make decisions. This broader approach is likely to offer students alternative opportunities to demonstrate their performance if they are disadvantaged by any one particular assessment in the programme (Gipps and Stobart, 2009).

It is also important that the format and design of different assessment instruments is informed by research on effective approaches for diverse student groups. The Australasian Curriculum, Assessment and Certification Authorities, for instance, recommended the distribution of "summaries of relevant research findings to item writers and test designers" (ACACA, 1995). In areas where there is limited research, for example inclusive assessment for students from cultural minorities, it is important for education systems to encourage the development of more research to extend the knowledge and evidence base.

Finally, equity in assessment is also about creating new attitudes, mentalities, and skills at every level of the educational system. This is a long-term process. As will be discussed below in the section on "Capacity", developing the competencies of teachers for inclusive assessment is key to avoiding bias in teacher-based assessment and improving equity in assessment.

Capacity

This third section discusses the competencies necessary to assess students, to benefit from assessment and to use the results of student assessment. It includes issues such as: the capacities students need to engage in and benefit from their assessment; the assessment competencies that teachers acquire in initial teacher education, professional development and moderation arrangements; and the expertise of the agencies involved in student assessment.

Student capacity for assessment

Traditionally, teachers have been regarded as responsible for establishing where learners are in their learning, where they are going, and what needs to be done to get them there. In recent years, there has been increasing focus on the role of the learner in assessment, not only as an active participant, but also as the critical connector between assessment and learning (Butler and Winne, 1995, in McDonald and Boud, 2003; Earl, 2003). While feedback by teachers and others provides information that can help students

improve, it is the students themselves who must make sense of that information, relate it to prior knowledge and take action to close gaps in their own learning. This is the regulatory process in metacognition. It occurs when students personally monitor what they are learning and use this monitoring to make adjustments in how they learn and understand.

Self- and peer-assessment are powerful processes that aim to enhance the role of learners in their own assessment. Self-assessment has been defined as "the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they met these criteria and standards" (Boud, 1986). Peer-assessment, on the other hand, involves students in assessing each others' work, again through reflection on goals and on what it means to achieve them. It may take place in pairs or in groups and has particular value in formative assessment as it allows students to use familiar language and ask each other questions that they may not dare to ask their teachers (EPPI, 2005). Such approaches to assessment can promote a greater sense of agency and responsibility of students in their own learning, and can help them engage in fruitful conversations about their learning leading to greater self-confidence, metacognitive monitoring skills and self-regulation skills, sometimes referred to as "assessment *as* learning" (Earl, 2003). A range of studies reviewed by Black and Wiliam (1998) report positive effects of student self-monitoring on the learning of different student groups (Sawyer et al., 1992; McCurdy and Shapiro, 1992; Masqud and Pillai, 1991; Merret and Merret, 1992).

Little information is available internationally regarding the extent to which students are engaged in their own assessment across countries. Some countries are beginning to implement policy frameworks that emphasise the importance of building the learners' own capacity for self-assessment and self-monitoring. In Ireland for example, the National Council for Curriculum and Assessment (NCCA) guidelines on assessment emphasise the importance of sharing learning goals with learners, helping learners to recognise the standards they are aiming for, involving learners in assessing their own learning, providing learners with feedback, communicating confidence to learners that every learner can improve, and adjusting teaching to take account of the results of assessment.

In Canada, several jurisdictions highlight the role of the learner in assessment within their curricula. In the Manitoba English Language Arts Curriculum for Years 5 to 8, for example, it is stated that:

> Modelling and encouraging metacognitive strategies helps students to understand, monitor, and direct their learning processes. Metacognitive questions such as, "What do you notice about your thinking?" and "How did you remember that information?" help students develop internal conversations and reflection about the learning process. When students have opportunities to reflect on their learning, especially with peers, they begin to develop self-assessment skills and want to take more responsibility for shaping and directing their own learning experiences. At times, students need quiet reflection. Whether alone or with others, students use reflection as a tool to consolidate what, how, and why they have learned, and to set goals for future learning.
>
> *(*Manitoba Education website, English Language Arts, Curriculum Documents, *www.edu.gov.mb.ca/k12/cur/ela/curdoc.html.)*

However, in many countries participating in the OECD Review, there were concerns that formative assessment approaches tended to remain teacher-centred rather than student-centred. While self- and peer-assessment are beginning to receive increasing attention in both policy and practice across countries, it appears that in many contexts

self-assessment is understood in a context of self-marking rather than reflection about learning. In several countries where OECD Country Reviews were organised, self- and peer- assessment practices were incipient, with little attention to sharing and co-constructing learning goals and criteria with students and involving students in reflecting on their progress and evaluating their learning outcomes.

To help students monitor their own learning, it is essential that they understand the learning goals and what they need to do in order to reach them (Sadler, 1989). Assessment schemes and purposes, as well as the specification of what will be assessed and against which criteria the judgement will be made, must be transparent to students (Ross et al., 1999). As students internalise the criteria for evaluating their work, they are better able to connect their performance with their preparation, and develop an internally oriented sense of self-efficacy (Stiggins, 2005). Teachers can use classroom assessment as the vehicle for helping students develop, practice, and become comfortable with reflection and with critical analysis of their own learning (Earl and Katz, 2008).

Teachers themselves also need to learn how to develop learner-centred teaching and assessment and how to introduce self-assessment practices into regular classroom activities. In a study conducted in Barbados, McDonald and Boud (2003) found positive effects of formal self-assessment training for teachers and students on student performance in external examinations. The study was conducted in the context of a large-scale introduction of self-assessment across a range of subjects, where teachers were trained in self-assessment practices and introduced these to a group of students preparing external examinations. The performance of students participating in this process was compared with that of a matched control group of students who were not given self-assessment training. The authors found a significant difference in performance with those trained in self-assessment outperforming the control group in each curriculum area.

Teacher capacity for assessment

How to best prepare teachers for their assessment responsibilities is the subject of debate in many countries because of the complexity of assessment and its integrated role with understanding teaching and learning. In the OECD's Teaching and Learning International Survey (TALIS), across the 23 participating countries, 15.7% of teachers indicated having "high professional development needs" in the area of student assessment practices in 2008. However, considerable differences can be observed across countries, with larger proportions of teachers (above 20%) in Brazil, Korea, Norway, Slovenia, Italy, Lithuania and Malaysia expressing high needs in this area (OECD, 2009).

Little comparable information is available internationally regarding the preparation and training that teachers receive to build their assessment capacities. In most countries, institutions responsible for teacher education are autonomous and define their own curricula, which naturally leads to variations across institutions regarding the content of initial teacher education. Information from education systems participating in the OECD Review provides some indications about the place given to assessment approaches within initial teacher education and professional development opportunities across countries.

Initial teacher education

Broadly, the information collected from education systems through the OECD Review appears to show that student assessment is given increasing attention in initial teacher education programmes internationally. Initial teacher education programmes vary

in the way that assessment is taught: *(i)* in a dedicated assessment course; *(ii)* within curriculum areas; *(iii)* theoretically; and *(iv)* practically.

In a few countries, such as Mexico, initial teacher education is centrally organised with a nationally uniform curriculum for teacher education university degrees. In Mexico, according to the national curricula for basic and lower secondary education, teachers should be trained to perform student assessment in the classroom during their initial education. Capacity to assess student learning is supposed to be developed through a course called "teaching planning and learning assessment", taught six hours per week in the sixth semester of the Bachelor's in Primary Education and four hours per week in the fourth semester of the Bachelor's in Lower Secondary Education. This course should include both assessment of learning processes and formative assessment practices (Santiago et al., 2012c).

Several countries report that while initial teacher education institutions are autonomous, they typically provide prospective teachers with basic knowledge and skills in student assessment approaches. This is the case for example in Australia and New Zealand (Nusche et al., 2012; Santiago et al., 2011). In Canada, many universities offer courses at the undergraduate level dealing with assessment practices. Some are compulsory courses for students working on becoming certified teachers, while others are offered on line for further professional development (Fournier and Mildon, forthcoming). In Finland, the assessment of students' progress and learning is taught in initial teacher education, as well as in vocational teacher education (Finnish Ministry of Education and Culture, forthcoming). In Korea, subjects such as educational evaluation, measurement and assessment of education, education research methodology, psychological examination, educational statistics and psychological measurement are provided as compulsory or optional courses for prospective teachers. Educational contents include the basic concept of educational evaluation, classification and types of assessment, principles and practice of test development, principles and planning of performance evaluation, sufficiency rating for test items, basic statistical analysis, results utilisation, and general classroom and student assessment (Lee et al., 2004, in Kim et al., 2010).

In a range of education systems, there have been central decrees or initiatives to restructure or regroup teacher education institutions, which allowed the central authorities to emphasise particular priority topics including student assessment. In the Flemish Community of Belgium, the Institutes for Initial Teacher Education were restructured by decree in 2006. According to the decree, the Institutes' curricula need to guide prospective teachers towards basic competencies, including skills for student assessment (Flemish Ministry of Education and Training, 2010). In the French Community of Belgium, initial teacher education has been redefined by decree in 2000 and now includes 30 units dedicated to assessment, differentiation of learning, identification of learning difficulties and definition of remedial strategies. Assessment approaches are also part of the pedagogical competencies to be covered throughout the initial teacher education for secondary teachers (Blondin and Giot, 2011). In Ireland, the content and duration of initial teacher education courses have been reconfigured as part of measures introduced under the National Literacy and Numeracy Strategy (Irish Department of Education and Skills, 2011) and the enactment of the Teaching Council Act 2001. The Teaching Council has also published criteria and guidelines which providers of initial teacher education are required to observe. In Norway, a framework plan for a new initial teacher education launched in 2010 provides guidelines regarding the development of teachers' assessment competencies. In particular, it requires that assessment for learning should be one of the competences that teachers have acquired upon graduation. In Sweden, a new initial

teacher education approach was to be implemented from July 2011. The new programmes contain specific goals related to assessment and marking. Assessment topics are also expected to be integrated into the didactics of every subject (Nusche et al., 2011b).

In Norway, the Directorate for Education and Training has also stimulated co-operation among teacher education institutions regarding the topic of student assessment. Until 2011, it funded the Norwegian Network for Student and Apprentice Assessment (NELVU), a network of teacher education institutions that aims to build capacity regarding student assessment within schools and university colleges. To this end, each teacher education institution has formed assessment experts within the institution to work with faculty on this particular topic. The focus was on all aspects of assessment literacy including the use of national test results, assessment for learning and different classroom assessment approaches. NELVU further aimed to stimulate research and development regarding assessment and co-operated with experts internationally, such as the Assessment Reform Group in England (Nusche et al., 2011a).

A recurrent concern across education systems that received an OECD Country Review was that there were variations in the degree to which different teacher education institutions made student assessment a priority in their programmes. In Australia, a survey of teachers revealed that "methods for assessing student learning and development" were among the areas of greatest need for professional development as identified by teachers (Santiago et al., 2011). In the Flemish Community of Belgium, the Inspectorate found in 2007 that starting teachers experienced that developing adaptive or alternative assessment approaches was one of their most difficult tasks, in spite of previous efforts to restructure teacher education with greater emphasis on such approaches (Flemish Ministry of Education and Training, 2010). In the Czech Republic, according to Santiago et al. (2012b), there is very little attention given to developing such assessment skills in initial teacher education programmes. Teachers are more familiar with using test score information for summative purposes and have not received significant training regarding the use of richer assessment tasks to inform their teaching. In Denmark, the OECD review team interviews revealed that pre-service teacher education programmes offered little training in student assessment for teacher candidates (Shewbridge et al., 2011).

Teacher professional learning

Teachers' professional learning regarding assessment is a career-long experience that needs to be sustained. In parallel to changes in initial teacher education, several countries have introduced a range of professional development and learning opportunities for teachers regarding assessment practice. Timperley (2011) describes the difference between professional development and professional learning. Over time, the term "professional development" has taken on connotations of provision of information to teachers in order to influence their practice whereas "professional learning" implies an internal process in which individuals create professional knowledge through interaction with this information in a way that challenges previous assumptions and creates new meanings.

The organisation of professional development courses regarding assessment is common practice in many countries and it appears essential in supporting a national focus on effective assessment practice. Gilmore (2008) makes a distinction between professional development programmes in which assessment is "foregrounded" (i.e. it is the main focus of the programme) and those programmes where assessment is "backgrounded", (i.e. the programme does not focus on assessment per se, but assessment

is an integral part of the programme). Many countries use a mix of both. In several settings, for example Hungary, Mexico and Norway, initiatives are also directed at school supervisors and/or school owners.

In Australia, most jurisdictions provide training to improve the competency of teachers to analyse and interpret student assessment data. For example, the Victorian Curriculum and Assessment Authority conducts in-service courses in schools around Victoria, to develop school leaders' and teachers' skills in interpreting the results of the national assessments and the Victorian Certificate of Education exam (Santiago et al., 2011). In the Flemish Community of Belgium, courses on assessment are an important part of the wide range of in-service training possibilities. It is common practice for schools to invite experts on various items (e.g. student assessment) to provide training opportunities for teachers (Flemish Ministry of Education and Training, 2010). In the French Community of Belgium, professional development on assessment is also available. The subjects dealt with in in-service training courses are determined based on needs identified by the *Commission de pilotage* and the Inspectorate (Blondin and Giot, 2011). In Korea, in-service training on educational assessment is provided as part of the national training framework. In recent years, local education offices and individual schools have also been adding new dimensions to the contents and methodology of such training (Kim et al., 2010).

In New Zealand, the Ministry of Education has initiated several major professional development programmes, which have been evaluated in terms of their impact on student learning, with promising results (Nusche et al., 2012). For example, "Assess to Learn (AtoL)" is a whole-school professional development programme that has been offered to primary and secondary schools since 2002. Schools can apply for participation in the programme and typically participate for two years. The annual budget for AtoL is NZD 3.17 million annually and currently involves 155 schools. The programme intends to support teachers in choosing adequate assessment tools and analysing assessment information so as to further advance student learning. A 2008 evaluation of the AtoL programme reported a significant impact of the programme on teacher professional practice and important improvements in student learning, especially for students with initially low achievement levels. Monitoring data showed that schools participating in AtoL had achieved up to 4.5 times greater shifts in writing achievements in Years 4 to 9 than the nationally expected rate of progress.

In some countries, the focus on professional development for student assessment has been considerably reinforced in recent years. In Hungary, for example, awareness raising campaigns with professional content on assessment were held in 2009 in every region. These occasions can also serve to strengthen the reputation and acceptance of assessments, and give professional impetus to make the use of results as diverse as possible. In Mexico, assessment-related topics are receiving increasing emphasis in the offerings available to teachers. While two years ago only two programmes were specifically focused on assessment issues, the 2011/12 catalogue includes over 30 programmes, among about 1 100 offerings. Most of them are targeted at school supervisors and focused on competencies-based assessment. Simultaneously, many subject-specific courses include new approaches, techniques and instruments for classroom-based assessment (Santiago et al., 2012c). In Norway, student assessment is also being highlighted as a key topic for the continuing professional development of school professionals and school owners. Since 2005, the Directorate for Education and Training has included student assessment as one of the annual priorities for continuing

professional development of teachers, school leaders and trainers of in-service training providers (Nusche et al., 2011a).

In several countries, professional development also takes place through moderation of teachers' assessment and marking. As explained above, moderation refers to quality assurance measures that seek to increase the consistency of marking, for example through teachers reviewing or cross-marking each other's assessments within a school or across schools or working together in groups to discuss assessment criteria and student performance levels. Moderation is a key strategy in validating consistency of teacher judgement and marking and it may occur within schools, between schools and across school sectors. At the same time, moderation also involves professional discussions between teachers about the quality of authentic pieces of student work and as such it has the potential to provide a powerful professional learning opportunity for teachers that they can relate closely to their classroom practices. It also contributes to improving teachers' professional judgements about student work and their developing a shared understanding of marking criteria or standards within schools and between schools (Timperley et al., 2008). This provides teachers with a chance to reflect on assessment in their subject, both on topics and criteria.

Finally, professional learning may also build on existing initiatives, provide opportunities for teachers and schools to network among each other or with assessment advisors and disseminate effective practice. In Norway, the Better Assessment Practices project (2007-09) supported a range of local projects to improve assessment practice in Norwegian schools. As a follow-up, the Assessment for Learning programme (2010-14) was implemented to support school projects and networks focusing particularly on formative assessment. There are also local initiatives in this area. The City of Oslo, for example, employs two "assessment advisors" that schools can invite to provide help regarding assessment (Nusche et al., 2011a). In Denmark, the availability of resource teachers at schools provides important support to teachers. Although assessment and evaluation advisors are few (reported in only 8% of schools) they have the potential to offer critical support to teachers whose initial training did not give particular emphasis to student assessment and evaluation (Shewbridge et al., 2011). In Canada, all Boards have created assessment divisions or sectors within their administrative structure and have assigned personnel to lead workshops, develop activities related to assessments, and to track, collate, analyse and distribute findings of district-wide assessments to their respective stakeholders (Fournier and Mildon, forthcoming).

However, while professional development opportunities in the area of student assessment appear to exist in most countries, there is often little information available at the central level regarding the extent to which teachers benefit from these offers and regarding the quality of available courses. In education systems where schools and teachers are free to determine the content of their professional development courses, it is often unclear to what extent teachers choose to improve their student assessment methods through such courses.

These concerns are linked to broader challenges in teacher professional development systems. While professional development is receiving increasing policy attention across countries, its provision is often patchy and uneven (OECD, 2005b). Teachers typically receive entitlements and incentives to engage in professional development, but these are not always matched by initiatives on the supply side. In some countries, teachers receive public funding only for courses offered by a few institutions, which can reduce the incentives for innovation and quality improvement (OECD, 2005b). Given the rapidly

changing demands being made of teachers in student assessment, countries may be faced with a lack of qualified trainers and well-designed training programmes in this area.

Guidelines and tools to support effective assessment practice

In addition to classic training courses, there are many other ways for education authorities to promote and support the strengthening of teachers' assessment approaches, for example through the development of guidance and support materials. In Denmark, for example, the electronic Evaluation portal provides a plethora of evaluation and assessment tools for use by teachers, schools and municipalities. This includes, for example, guidance materials on how to assess students in the final examinations in Form 9 prepared by the examiners and subject advisors of the final examinations, including advice and ideas for classroom teaching and criteria for student evaluation and assessment in classroom activities (Shewbridge et al., 2011).

In Ireland, the National Council for Curriculum and Assessment (NCCA) guidelines on assessment provide detailed, practical guidance to teachers on how to use a range of assessment approaches in order to obtain a full picture of a child's achievement. They also provide guidance to schools on the development and implementation of an assessment policy. The NCCA has also contributed to the development of expertise in relation to formative assessment through its curriculum development projects with schools and designed materials that are intended to support teachers and schools in expanding their assessment toolkit for classroom-based assessments (Irish Department of Education and Skills, 2012).

In several countries, guidance materials have been developed with the aim of strengthening teachers' understanding of standards and criteria used in central assessments and examinations. In the Slovak Republic, for example, following the organisation of external assessments, the National Institute for Certified Educational Measurements (NÚCEM) publishes analytical reports for teachers with recommendations for improving the quality of education according to measured findings along with methodological guidelines and collections of tasks. NÚCEM also organises expert seminars for teachers to provide participants with feedback on test results (Hajdúková et al., forthcoming). In Ireland, the availability of marking schemes from State examinations combined with the publication of Chief Examiners' reports supports greater understanding among teachers of expected standards and criteria.

Involving teachers in marking central assessments and examinations

In several countries, professional learning is organised around central assessment or examinations. Hiring teachers in order to correct and mark central assessments may contribute substantially to their understanding of expected standards and criteria. In Canada, teachers are involved in the development of regional, provincial and national tests and the associated materials such as scoring guides, rubrics and the choosing of exemplars for use in scoring. Teachers are involved in all stages of the administration process (including field testing) and they are also trained for the scoring of completed assessments, using the prepared rubrics and exemplars. The result has been the accumulation of experienced test developers who will enhance their own teaching practice and may bring to their school communities a rich source of leadership in assessment.

In New Zealand, teacher professional development related to effective assessment occurs via their strong involvement in scoring student work for the National Education Monitoring Project (NEMP) and for the National Certificate of Educational Achievement

(NCEA). For the National Education Monitoring Project, about 100 teachers are freed from their teaching responsibilities each year to conduct the assessments. They receive one week of training and then administer the tasks over a period of five weeks. The intention is to ground the assessment practice in sound teaching practice and to build and strengthen teacher assessment capacity.

In Norway, professional development also takes place around teachers' marking of central examinations and in moderated marking of oral examinations. In addition, some school owners further support moderated marking processes. In 2010, the municipality of Oslo launched a pilot study in lower secondary education, where they invited all schools to implement a mock exam. The municipality invited 60 teachers from 35 schools to come together to mark the examinations in a moderated marking process in collaboration with expert teachers (who had been part of the national marking process). This provided an opportunity for teachers from Oslo schools to engage in discussion about the meaning of marking criteria in relation to examples of student work.

Concerns about fragmented approaches to teacher learning in assessment

In spite of remarkable efforts across countries to ensure that teachers acquire basic assessment literacy through initial teacher education programmes and continuing professional learning, countries also report that more needs to be done to help teachers develop the competencies necessary for effective assessment. In all countries, there are indications of some inadequacies in teachers' preparation for student assessment. New teachers sometimes need substantial support to develop classroom assessment techniques, reporting schemes and moderation processes, which in some cases are not priority areas during teacher education programmes. More experienced teachers may lack access to continuing professional development supporting them in the use of most recent assessment technology and staying up to date on emerging research on effective assessment approaches. Experience from across OECD countries reveals that the following areas require particular attention:

- **Assessing key competencies**. While internal assessment provides opportunities for diverse and innovative assessment approaches, its validity depends to a large extent on the assessment opportunities provided by individual teachers (Harlen, 2007). Indeed, reviews of research on teacher-based assessment note that teacher-made assessments often focus on content knowledge and may be no more diverse or innovative than external tests (e.g. Crooks, 1988; Black, 1993; Black and Wiliam, 1998; Harlen, 2007). Hence, it is important to ensure that teachers receive adequate training to assess a broader range of skills and competencies (Somerset, 1996).
- **Effective marking and reporting**. There are sometimes large variations in the ways teachers assess students and set marks. This is compounded in situations where there are no central marking criteria and where there is no guarantee that teachers engage in discussion or moderation within or across schools. In the case of summative assessments that carry high stakes for students, this poses important challenges to the fairness of assessment and marking. It is important to provide focussed training on how to make summative judgements on student performance in relation to central curriculum goals or standards. Such training should include how to identify valid evidence and how to apply scoring rubrics and marking criteria to very different types of evidence of student learning (Harlen, 2004).

- **Effective formative assessment**. Ongoing attention to teacher training in formative assessment is also vital. An important priority is to develop teachers' capacity to interpret student assessment data, including that generated by standardised tests, for the improvement of classroom instruction. To become assessment literate, teachers need to be aware of the different factors that may influence the validity and reliability of results and develop capacity to make sense of data, identify appropriate actions and track progress. Other key areas of training in formative assessment are to help teachers provide effective feedback to students and to fully engage students in their own assessment (Looney, 2011a; Earl and Fullan, 2003).
- **Inclusive and fair assessment of diverse students**. Several studies report that there are risks of conscious or unconscious bias in teacher-based assessment (Crooks, 2004; Harlen, 2007). Bias in teachers' assessment may be related to teachers' prior knowledge of student characteristics such as behaviour, gender, special educational needs, immigrant background, first language, overall academic achievement or verbal ability (Harlen, 2004; Gipps and Stobart, 2004). The key challenge is to ensure that rich assessment opportunities are systematically offered to all students regardless of such characteristics. It is important that dimensions of inclusive and equitable assessment are covered in both initial education and professional development to help teachers adapt their assessment approaches to the diversity of student needs in their classrooms.

What is missing in many countries is a *strategic* approach to teacher learning in assessment. While teachers may learn bits and pieces about effective assessment at various stages of their career, the offer of professional learning in this area appears fragmented and limited in scope. The vision of assessment communicated in professional development courses is not always well aligned to the way assessment is covered in initial teacher education. Also, the different professional development programmes regarding assessment are typically run independently of each other, often without recognition of overlaps and synergies. The development of teacher standards or professional profiles – and the inclusion of assessment as an important teacher competency in these standards – can help provide purpose and structure for professional development at different stages of the teaching career (Chapter 5). There is a role to play for the central authorities to ensure alignment of publicly funded professional development courses so that they foster a coherent understanding of the interrelations between curriculum, teaching, learning and assessment (Absolum et al., 2009).

Central capacity for assessment development and implementation

In all education systems, the educational administration and/or a range of central agencies are involved in student assessment in various ways. As explained in Chapter 3, several countries have created specialised assessment agencies in recent years, reflecting the growing importance of educational measurement and the increasing complexity of evaluation and assessment frameworks across countries. In Austria, for example, the Federal Institute for Education Research, Innovation and Development of the Austrian School System (BIFIE) was created in 2008 to develop, implement and monitor education standards. Similarly, in the Slovak Republic, the National Institute of Certified Measurement (NÚCEM) was created by the Slovak Ministry of Education in 2008 to develop, administer and oversee all national tests and assessments. Other countries have more long-standing agencies in charge of overseeing and developing the education system's assessment strategy. Expertise regarding education standards and assessment is

typically also concentrated in specific departments of the Ministries of education, Inspectorates and education review bodies, curriculum organisations, quality assurance agencies and examinations and qualifications authorities (for a detailed overview, see Chapter 3).

Student assessment is a highly technical matter and the design and implementation of standardised assessments requires expert capacity which takes time to be developed. In several education systems where OECD Country Reviews took place, there were concerns about the lack of expertise at the central level in the area of student assessment. Such expertise is particularly important when education systems are developing and introducing large-scale assessments. Considerable investment is needed to develop capacity and expertise in standardised test development, including areas such as educational measurement and assessment design.

In addition, a deep understanding about the role of assessment in the improvement of policy and practice is essential to develop policy, support the development of central assessment tools and ensure that assessment results are used appropriately for monitoring and support of schools. Central agencies can build such capacity by engaging with each other and with the assessment community more widely, both nationally and internationally. For example, the establishment of independent working groups from a range of sectors and organisations in education can help accompany the development of central assessment strategies, by debating the central assessment instruments, monitoring their implementation and conducting impact evaluations. Such a group should have the remit and expertise to investigate and make recommendations that centrally mandated assessments are valid and reliable.

Some tools are also available internationally to help governments and central agencies review and further develop student assessment systems. The World Bank, for example, has developed a set of tools called SABER (Systems Approach for Better Education Results, available at *www.worldbank.org/education/saber*) to support countries in systematically reviewing the strengths and weaknesses of their student assessment systems. These tools comprise a set of standardised questionnaires and rubrics for collecting and evaluating data on three types of assessment (classroom assessment, examinations and large-scale system-level assessment). Policy makers and central agencies can use these tools to benchmark their system-level policies and practices and plan for further development. However, it is important to note that additional tools will be necessary to determine actual assessment practices implemented in classrooms across a given country (Clarke, 2012).

It is also the role of the central administration and agencies to establish and share a vision for assessment in the school system. This requires the development of a comprehensive communication strategy to engage stakeholders at all levels in dialogue regarding effective assessment. Leadership and communication capacity required at the central level in order to build a common understanding of the different purposes and uses of assessment in education. In Norway, for example, the Directorate for Education and Training has created a department on internal governance to enhance continuous reflection about the uptake and impact of new initiatives for quality improvement in the education sector. It has also introduced a professional development programme to build leadership among its staff and to enhance effective goal-setting and strategy development within the Directorate itself (Nusche et al., 2011a).

Reporting and use of results

This section is concerned with how assessment results are reported and used for both summative and formative purposes. It describes standards of quality and reporting formats used in different contexts, reviews the legal frameworks in place to regulate reporting of results and discusses the ways in which assessment results are used in different contexts to record information, provide feedback to students and make decisions about their further educational trajectory. The ways in which assessment results are recorded and utilised are essential to the distinction between summative and formative assessment.

Regular summative reporting helps communicate summary statements of achievement to students and parents. As such, it can help reach out to parents and involve them in supporting their children's learning. Records of student achievement can also help teachers communicate about student achievement within a school and make decisions about targeted support. They can contribute to ease transitions when students are changing schools and to ensure consistency between different levels of education. In many countries, such records are also used to make decisions regarding the student's educational trajectory, for example for decisions regarding school entry, year repetition, transfer and ability grouping.

The formative collection and use of results, on the other hand, is embedded in the process of teaching and learning itself. The distinguishing feature of formative assessment in all countries is that the information is used to make improvements (Bloom, 1968; Scriven, 1967). But the way in which information is used and the timescale for decisions may vary across teachers, schools and education systems. Wiliam (2006) distinguishes between long-, medium, and short-cycle formative assessment. According to Wiliam, long-cycle formative assessment occurs across marking periods, semesters or even years (four weeks to a year or more); medium-cycle formative assessment occurs within and between teaching units (three days to four weeks); and a short-cycle formative assessment occurs within and between lessons (five seconds to two days).

Overall, the utility and sound use of data, of course, depends on teachers' assessment literacy and ability to appropriately integrate assessment data and learning in classroom instruction, including the appropriate use of standardised tests. This means that teachers and school leaders need to continually develop their capacity to collect and report on student assessment to students, parents and external partners in effective ways without oversimplifying the complex issues involved in student learning.

Standards of quality and reporting formats

The interpretation of any kind of measurement depends on the standards of quality that are applied. A student's performance may be measured in three different ways:

- **Criterion-referenced** (performance in relation to established standards or criteria). Criterion-referenced assessments are used to make judgements about absolute levels of performance. Such assessments may set benchmarks for what constitutes "mastery" or "high performance" and/or determine minimum standards that should be achieved by every student.
- **Norm-referenced** (performance in relation to a defined group). Norm-referenced assessments classify students based on a comparison among them. The results of norm-referenced assessments have meaning only in comparison with the results of

other students. They do not reflect their proficiency in relation to absolute standards, but in relative terms.

- **Self-referenced** or **ipsative** (change in performance over time). Self-referenced assessments are generally used formatively by teachers to track the growth and progress of individual students over time.

Countries tend to use a mix of different quality standards and reporting formats depending on the specific purpose of each assessment. It can be argued that criterion-referenced assessments are more useful to inform future teaching and learning because they measure student performance against specific learning goals. Norm-referenced assessments, on the other hand, may be more useful for the purpose of student selection (e.g. for university admission) because they allow to compare students among each other (Looney, 2011a). Criterion-referenced assessments tend to be more common for the purpose of summative examinations in lower secondary education (OECD, 2011). Of 13 countries for which information was available, only 2 were using norm-referenced assessments, while 10 were using criterion-referenced assessments. In the United States, both criterion-referenced and norm-referenced assessments were allowed with decisions taken at the school level. A similar picture can be observed in upper secondary education: of 19 OECD countries for which information was available, only 5 were using norm-referenced central examinations while 14 countries were using criterion-referenced examinations (OECD, 2011).

For assessment systems to be well aligned with curriculum goals, teachers are generally expected to assess students in relation to central curriculum goals and standards (Table 4.1). However, the OECD Country Reviews indicate that where assessment standards and criteria are not clearly formulated for different subjects and year levels, teachers often use their own personal reference points, based on their experience and school-based expectations. Teachers' classroom-based assessments often are a mixture of norm-referenced (in relation to other students), content-referenced (in relation to what they taught) and self-referenced (progress of individual students). In some countries, there were concerns that teachers tended to give norm-referenced marks by comparing a student with other students in a class. Such comparative assessments are no longer the most appropriate nor informative frame of reference for teachers to use in systems that have established a focus on learning outcomes and standards of expected achievements.

Central frameworks for summative reporting

As shown in Tables 4.A1.3, 4.A2.3a and 4.A2.3b (Annexes 4.A1 and 4.A2), almost all education systems for which information is available have policy frameworks that regulate the reporting of summative results. Such reporting frameworks tend to be developed at the central/state level and in most education systems they are compulsory for all schools[10]. The frameworks typically determine the frequency of summative reporting and the type of information to be provided to students and their parents. In Australia, for example, summative assessment practices are set at the state/territory level but there is a national requirement to report these in a nationally consistent way. Education systems vary regarding the frequency of formal reporting. In most systems, student results are reported in writing on average twice a year, but this may vary depending on the level of education.

The type of information provided in regular summative reports also varies across countries and levels of education. While the use of numerical, alphabetical or descriptive (e.g. excellent, good, satisfactory, unsatisfactory) marks is common at all levels of

education, it becomes more prevalent at the secondary level. In primary education, students are not awarded marks in Denmark, Iceland, New Zealand, Norway, Poland (Years 1-3 only), Slovenia (Years 1-3 only) and Sweden (Years 1-5 only). These countries rely instead on regular qualitative reporting, for example in the form of a summary of the student's strengths and weaknesses. Australia, Austria (Years 1-2 only), Finland, France, Ireland and Israel tend to use a mix of qualitative assessments and marks in primary education, whereas Hungary, Italy, Mexico, the Netherlands, Poland (Years 4-6), Slovenia (Years 4-6) and the Slovak Republic rely primarily on numerical marks for formal reporting. In Canada, requirements vary across provinces, and in Estonia, schools can decide which type of marking they use. At the lower and upper secondary level, all countries provide summary marks or ratings to students in regular reports. A number of countries, including Australia (general education only), France, Iceland (lower secondary only), Israel and Sweden (for students who fail in lower secondary only) complement these with qualitative written assessments.

In addition to requirements for regular written reporting, countries also frequently have central requirements for teachers to hold formal summative feedback meetings with students and their parents. Again, the most common periodicity for such meetings is twice a year, but there are variations across countries and levels of education. In France and Ireland, parents can request additional meetings beyond the mandatory meetings. In Spain, summative feedback meetings take place only at the request of parents. Denmark and Iceland have no particular requirements for such meetings in secondary education, and the Flemish and French Communities of Belgium, Finland, Hungary, Italy, Korea, New Zealand and Poland do not have such requirements at any level of education.

Finally, there are also variations across countries regarding the type of certification students receive at the end of key stages of education. At the end of primary education, just over half of the education systems for which information is available provide students with a formal certificate of learning (Tables 4.A1.3, 4.A2.3a and 4.A2.3b, Annexes 4.A1 and 4.A2). In most education systems where formal certificates are awarded, they contain a mark in each relevant subject at the end of primary education. In Finland, students usually also receive such report cards even though the national authorities do not require numerical marks before the end of Year 8. In France and Slovenia the subject marks are complemented by the overall average and a summary of the student's strengths and weaknesses (up to Year 3 in the case of Slovenia). In Poland, marks for behaviour, additional coursework and special achievements are also included. In Mexico, the formal certificate also includes Pass/Fail information and an overall mark. In Korea, the certificate only provides Pass/Fail information. In Northern Ireland in the United Kingdom, formal certificates of learning at ISCED levels 1 and 2 report the student's level of progression achieved in Communication and Using Maths (and in Using ICT from 2013/14).

At the lower secondary level, only the French Community of Belgium, Canada, the Czech Republic, the Slovak Republic and New Zealand do not award formal certificates of learning. In the majority of education systems, students receive a mark in each of the concerned subjects. In Finland, students receive an overall mark in addition to marks in each subject. Students receive Pass/Fail information in Korea, Mexico (in addition to an overall mark) and the Netherlands (in addition to a mark in each of the concerned subjects). In Denmark, in addition to subject marks, there is also a written assessment of a student project conducted in Year 9. Students also receive additional information in France (overall average and a summary of strengths and weaknesses), Iceland (qualitative assessments), Italy (overall mark based on an average of all tests including the national examination) and Poland (marks for behaviour, additional coursework and special

achievements). Some countries deliver a different type of certificate in pre-vocational and vocational programmes, such as the statement of attainment of competencies in Australia.

At the end of upper secondary education, all OECD education systems deliver a form of written certification, graduation report or diploma to students having completed the programme. In most systems, this takes the form of marks in each of the concerned subjects, overall marks or certain additional information.

Reported concerns about marking schemes

Marks awarded to students play an important role in all education systems. They may inform students and parents in a concise form about performance and return to efforts in learning processes. They can also potentially improve performance if they help to raise students' aspirations and convey that there are high returns to effort. Marks constitute an important source of information, as students rely on them to develop their expectations for further learning and enrolment in higher levels of education (OECD, 2012b). On the other hand, there are also risks that written marks may discourage effort and motivation if the information hurts self-confidence or convey to a student that returns to effort are low (Sjögren, 2009).

Also, across the countries participating in the OECD Review, there were a number of concerns regarding the ways in which marks were constructed and awarded to students. While most countries regulate the use of a particular marking scale (e.g. A-E or 1-10), especially in secondary education, this does not mean that the meaning of a particular mark is necessarily equivalent across schools. Even if schools use the same marking scale, they may have different marking criteria. It is difficult to ensure that the marks awarded in one school align with similar marks in another school. In addition, the same teacher will not necessarily be consistent in the application of criteria across students and over time. Such inequity in marking becomes problematic when a student moves to another school, or when marks are used for high-stakes decisions.

In many countries, teachers reported that overall marks were made up by assigning points to students across a range of elements including, for example, attendance, participation, homework completion, test performance, presentations, teamwork, neatness and discipline. However, the weight assigned to each of these elements was often not made explicit and varied across teachers, subjects and schools in the absence of rubrics specifying the meaning of points for each of these activities. Establishing marks by averaging points across a range of tasks and behaviours is likely to result in a kind of grand number with no clear significance for students (Ravela, 2009). Hence, it becomes difficult for students to understand what is expected of them, and obtaining a high mark may become the main objective for them with potentially distorting effects on learning efforts (Shepard, 2006). Parents may also become more concerned with the marks than about the actual learning progress of their children.

In several countries, there were concerns about a conflation of marks for performance and marks for behaviour. The practice of combining the assessment of behaviour with the assessment of actual achievement risks undermining the role of marks as indicators of performance. It may provide incentives for students to simulate effort as they learn how to behave with each teacher to make a good impression. Marks may also risk becoming a disciplinary instrument for teachers to control student behaviour rather than a tool to inform about learning and create a motivating learning environment (Shepard, 2006). While behaviour, effort and motivation are undeniably very important factors influencing student achievement, it would be preferable to report on these aspects separately. This

would allow communicating more accurate information about the student as a complex learner and provide indications about how particular learning needs can best be addressed. Education policy makers should consider providing guidance regarding appropriate reporting about student learning (see Box 4.11 for examples from Canada).

Box 4.11 Canada: Policies on reporting student performance

Similar to the policies of other jurisdictions in Canada, **Halifax** teachers are asked to develop assessment plans that are aligned with the Halifax Regional School Board of Nova Scotia *School Plan for Communicating Student Learning*. These plans must outline the purposes of the assessment mechanisms in light of the intended audiences and must be founded on individual learning practices accurately reflecting the learning outcomes as defined by the provincial curriculum and/or individual program plans. The policy indicates that comparisons of performance between students are to be avoided and that evaluation should not be based on measures such as students' social development and work habits, bonus points, student absence, missed/late assignments, group scores, neatness, etc. In addition, if report card marks are used, these should emphasise evidence collected from students' summative assessments. The policy also states that all actors (teachers and students) are called to interact continuously during the formative assessment process in order that information used to judge the progress leads to an understanding about the desired outcome and the criteria used to determine whether these outcomes have been successfully completed.

As for other Canadian jurisdictions, **Nova Central** School District's policy and procedures clearly indicates that a mark is not a simple reflection of a set of averages on a series of assignments but that it should be linked to a combination of evidences gathered from a variety of sources that students know and recognise. It is worth mentioning that teachers are directed not to look at the cumulative measures of achievement over time but that the final summative judgement be based on the student's most recent demonstration of his achievement of the desired outcome.

Source: Fournier and Mildon (forthcoming).

Reported concerns about communication with parents

Despite the existence of basic requirements regarding marking and certificates, in several countries there are no clear rules on how teachers should communicate assessment results to students and parents, and regular reporting practices are highly variable. In some settings, parents receive very limited feedback regarding their children's performance and progress. Throughout the OECD Country Review visits, parent representatives in different countries indicated that insufficient information was provided to them regarding the progress and identified needs for improvement of their children. In this context, several countries are working towards improving their reporting formats and harmonising reporting approaches across schools (for an example from Australia, see Box 4.12 below).

Absolum et al. (2009) criticise the very notion of "reporting" to parents as implying a power relationship that may inhibit meaningful partnership and dialogue. The term in fact suggests that parents are passive recipients of information rather than active partners of schools and supporters of their children's learning. The authors advocate for a mutual exchange of information between schools and homes where parents also share their understanding of their child's learning with the school. To establish such reciprocity, schools need to ensure that the information they provide is clear and easily understandable. Useful information, beyond simple marks, would include details about

students' progress, strengths, areas of need or concern, recommendations for further learning and illustrative examples.

Box 4.12 Innovations in the reporting system in Australia

In **Australia**, parental feedback at national, state and territory forums suggested that parents were confused by the different reporting scales and mechanisms used across schools. In response to these concerns, the Australian Government brought a degree of standardisation to teachers' judgements by requiring in 2005 that each state and territory adopt a common five-point scale as a condition for federal funding. At each year level from Year 1 to Year 10, teachers have to report students' achievements to parents using an A-E (or equivalent) framework. Defined specifically by each state and territory, generally the points on the A-E scale represent: A means well above standard; B means above standard; C means student at expected standard at time of report, on track to proficiency; D means below standard; E means well below standard.

A-E ratings are intended to assess the full range of learning expected of students. States and territories have developed guidelines and definitions for each of the A-E levels, variously labelled as letters (A-E) or descriptive categories (e.g. advanced – developing). States and territories vary in the specificity of the definitions and guidance they provide to support consistent judgements across teachers and schools. For example, Victoria provides teachers with detailed standards (*Victorian Essential Learning Standards*) co-ordinated with expected progression points, assessment maps and assessment modules to gauge student progress. Reporting software enables teachers to enter assessment scores and other ratings for components of each standard and the system automatically aggregates these scores into overall ratings for each student. To support consistency, Victoria also examines the relationship between the distribution of students' A-E ratings and NAPLAN results. There is a proposal to link A-E standards to the Australian Curriculum. The work has started in 2011 on a common approach to the achievement standards across states and territories including trialling and validation. In future, part of the work to align A-E with the Australian Curriculum will involve national agreement on definitions.

Source: Santiago et al. (2011).

Summative use of results

Cautions regarding potential misuse of summative assessment results

The results of summative assessment may be used by a range of stakeholders within and beyond schools, many of whom will have little or no training regarding measurement of student learning. Not all users of assessment results will be familiar with the intended uses of particular assessments, the content of the assessments, the evidence concerning the validity of inferences, and the characteristics of the test-takers. All of these information gaps may increase the likelihood of misuses of test results (Camara, n.d.). There is a large body of literature about standards for appropriate uses of assessment results. The *Standards for Educational and Psychological Testing* developed by AERA, APA and NCME (1999) are probably the most widely used reference in this domain.

Respecting a number of principles regarding the appropriate use of assessment results is essential to ensure responsible professional practice in educational assessment. Such principles include providing multiple opportunities for students to take alternate forms of an assessment, with opportunity for further learning between assessment administrations; drawing from multiple sources of information to enhance the validity of decisions being

made based on assessment results; considering the substitution of alternative measures for test scores, especially when tests are likely to give a deceptively low indication of a student's achievement because of test anxiety or disabilities that may reduce the validity of standardised test results; and ensuring that students have indeed had opportunities to learn the material that they are being assessed on (AERA, APA and NCME, 1999; Baker and Linn, 2004; Camara, n.d.).

Uses of summative assessment results across countries

Across OECD countries, the results of summative assessments are typically used for a wide range of decisions, including both within-school decisions and decisions beyond the school such as transfer to another school, transitions to higher levels of education and entry to the labour market. The remainder of this section provides an overview of country practices.

School-internal use of results

The primary function of summative assessment appears to be to keep records within schools and report to students, parents and other teachers. In most countries, summative assessment is conducted on a regular basis to collect information on student progress and check up on what students have retained and learned from a series of teaching units over a period of time (Harlen, 2007). This information is typically summarised in the form of marks and transcripts that students obtain at the end of course or year level. The aim most often is to provide a record of progress in learning to students, parents, other teachers and school leaders. To manage student assessment information on a longitudinal basis, some countries use computer applications to store student performance results, along with other information such as socio-demographic background and attendance.

Decisions on school admission, tracking and transfer

Schools may use summative assessments to decide on the admittance of students. This is more often the case in secondary than in primary schools. According to PISA 2009, on average across OECD countries, 36% of 15-year-old students are enrolled in schools whose principals reported that their schools are highly selective. In 10 of the 34 OECD countries, more than half of all 15-year-old students attend schools that always consider recommendations from feeder schools or academic transcripts when making admission decisions. Of these countries, in the Netherlands, Japan and Hungary, according to information provided by principals at the lower secondary level surveyed in PISA, more than 85% of students are selected for schools on the basis of academic records or recommendations. By contrast, less than 10% attend academically selective schools in Portugal, Spain, Iceland, Sweden, Finland, Denmark, Norway and Greece (OECD, 2010a).

Tracking into different school types

In a few OECD education systems, results are also used for tracking students into different school types at the end of primary education. In these cases, the assessment results of students in primary education have an effect on their transition from primary to secondary schools. In Europe, this is the case in Austria, Germany, Luxembourg and the Netherlands, and to some degree in the Flemish and French Communities of Belgium and Switzerland. In Austria, students need to have completed the fourth year of primary school with good marks in German and mathematics to be admitted to academic

secondary schools (*Gymnasium*). Students who are not admitted may take an entrance examination set by the *Gymnasium*. In Germany, after four years of primary education, primary schools give a recommendation on the type of lower secondary education the student should attend. Depending on the *Land*, the final decision is taken by the parents, the secondary school or the school authority. In Luxembourg, primary schools give a recommendation at the end of Year 6 in primary school, but students who are not admitted can take a national entrance examination. In the Netherlands, students' choice of secondary school is conditioned by their primary school's advice regarding the most appropriate programme of secondary education. This advice is based on the judgement of primary school teachers about the capacity of the student and on the student's results in a standardised but non-compulsory test in the final (sixth) year of primary education (Eurydice, 2009b).

Decisions on transferring students to another school

Assessment results may also be used to inform student transfers from one school to another. Transferring students to other schools because of low academic achievement, behavioural problems or special learning needs is a way for schools to reduce the heterogeneity of the learning environment. On average across OECD countries, at the lower secondary level, 18% of students attend a school in which school principals reported that the school would likely transfer students with low achievement, behavioural problems or special learning needs. Yet transfer policies vary across countries: in Iceland, Ireland, Portugal, Norway, Finland, the United Kingdom, New Zealand, Australia and Sweden less than 5% of students attend schools whose school principals reported that the school would likely transfer students for these reasons. By contrast, in Luxembourg, Austria, Belgium, Greece and Turkey, around one-third or more of students attend a school whose principal reported that students with low achievement, behavioural problems or special learning needs will "very likely" be transferred out of the school (OECD, 2010a).

Ability grouping within schools

Within schools, student assessment results may also be used to "stream" students into groups of similar ability levels. In PISA 2009, principals of lower secondary schools were asked to report whether students were grouped by ability into different classes or within a class, and whether these groupings were made in all or only selected subjects. Schools in most OECD countries use student assessment results in order to group students by ability within classrooms or between classrooms. On average across OECD countries, 13% of students are in schools whose principals reported that students are grouped by ability in all subjects, 55% are in schools whose principals reported that students are grouped by ability in some subjects and 32% are in schools whose principals reported that there was no ability grouping (OECD, 2010a).

Year repetition

In some countries, summative assessment results are used to inform decisions about year repetition (see also Chapter 3). This means that the results may be used to identify under-achieving pupils and decide on their progression from one year to the next. In many OECD countries, schools can decide to make students repeat a year if the assessment shows that they have not acquired adequate mastery of the curriculum at the end of a school year. However, there are large variations in country practices regarding the extent to which this possibility is used. In PISA 2009, on average 15% of 15-year-old students reported that they had repeated a year at least once: 7% had repeated a year in

primary school, 6% had repeated a year in lower secondary school and 2% had repeated an upper secondary Year. There are, however, marked differences across countries. While year repetition is non-existent in Korea, Japan and Norway, over 25% of 15-year-old students in France, Luxembourg, Spain, Portugal, Belgium and the Netherlands reported having repeated a year during their school trajectory in primary and secondary education (OECD, 2010a).

Decisions beyond school

In most OECD countries, a school leaving certificate is awarded to students after successful completion of compulsory education. In many OECD education systems, this certificate is at least partly based on results achieved in a final examination. But in some systems including Austria (up to 2013/14), the Flemish Community of Belgium, Finland, most German *Länder*, Hungary, Luxembourg and Turkey, the school leaving certificate is only based on teacher-assigned marks and work over the year. In several countries, the certificates are awarded on the basis of a combination between school-based marks and tests and external standardised exams (Eurydice, 2009b).

In all OECD countries, students receive a certificate after successful completion of upper secondary education. This certificate is generally a minimum requirement for admission to tertiary education. Only in Spain and Turkey the certificate is awarded only on the basis of continuous school-based assessment. In all other countries, certification is based at least partly on some kind of a final examination, often a combination of internal and external (or externally verified) exams. In France, Ireland and Slovenia, certification is based entirely on an external final exam (Eurydice, 2009b, updated with information from OECD Country Reviews).

In some education systems, the successful graduation from upper secondary school is a minimum requirement for entrance into higher education, for example in the Flemish Community of Belgium, France, Germany and the Netherlands. This does not mean that students are necessarily automatically admitted to higher education, but institutions of higher education and faculties may have their own supplementary entrance exams. Other countries such as Ireland, England, Korea and Australia directly use assessment for qualification and certification in upper secondary education for selection into higher education. In a further group of countries, upper secondary certification may serve a selection purpose only for subjects which are on high demand (Dufaux, 2012).

Even though the labour market does not have any minimum entrance requirements, students who do not choose to continue in higher education or another educational programme, but to apply for a job may undergo a similar process of selection. If the certificate of upper secondary education is perceived as a trustworthy institution for the communication of students' skills, employers may strongly base their selection of employees on the information provided through the certificate. Thus, certificates may have a strong screening or signalling function for employers. Reducing information asymmetry, employers may use certificates as a reliable tool to decide whether the applicant matches the requirements (Dufaux, 2012).

Formative use of results

As described above, the majority of countries participating in the OECD Review now have policy frameworks in place to support and promote formative assessment. At the same time, little comparable information is available about the way teachers actually document and use formative assessment results in practice.

Several countries have implemented measures for long- and medium-cycle formative assessment. This includes, for example, the use of Individual Development Plans (IDPs) in compulsory schools in Sweden. School leaders are required to set out the shape for the IDP. The Plans are to include an assessment of the student's current performance levels in relation to learning goals set in the curriculum and syllabi, and the focus is on steps that the student should take to reach those goals. Whether to include additional information, such as the student's more general development (e.g. the student's ability to take on responsibility, their social skills, and so on) is up to the school leader. For students who are experiencing difficulty, schools are required to document plans as to how they will help students achieve goals. The goals determined in IDPs are also used for student self-assessment in which students are asked to rate their own progress and performance. The IDP ensures that both teachers and students are focused on identifying individual learning goals, and developing strategies to address any shortcomings (Nusche et al., 2011b).

In Denmark, mandatory Individual Student Plans (ISPs) were introduced in 2006 to document student learning progress. According to Shewbridge et al. (2011), the ISPs contribute to formalising Danish assessment practice by documenting students' learning progress for dialogue with key stakeholders. They emphasise the student's future learning rather than summative learning outcomes. Official evaluations, strong support from national level parent organisations and student associations (see Danish Ministry of Education and Rambøll, 2011) and stakeholder feedback during the OECD Country Review confirm that the ISPs are well received by parents and teachers. In short, parents appreciate a written summary of their child's progress because they feel that they are better prepared for their meeting with teachers. Teachers perceive benefit in transferring documented information on student achievement to subsequent teachers and as such ISPs play a crucial role in tracking individual students' developmental growth over time. Teachers recognise the role of ISPs in easing communication with parents. The added workload ISPs entail for teachers is a bone of contention, but there is a current pilot to allow educators more flexibility in determining and prioritising the content of ISPs. Depending on the evaluation of this pilot, this may lead to a modified approach to drawing up ISPs (Shewbridge et al., 2011).

In the French Community of Belgium, the monitoring of students' progress towards core competencies is organised through individual learning plans (*Plans Individuels d'Apprentissage, PIA*). These plans are obligatory for students in special needs education, students with special learning needs who are integrated in mainstream education, as well as certain students facing important challenges in secondary education. The plan lists both subject-specific and transversal competencies to be acquired by students during a specified period. It is personalised in line with the potential and needs of each individual students and is regularly adjusted by the class council, based on observations provided by teachers and the school's guidance centre. The plans are used for formative purposes only, and are designed as much as possible in collaboration with the concerned students and their parents.

In Ireland, the National Council for Curriculum and Assessment (NCCA)'s assessment guidelines suggest that three records be maintained by primary schools: the teacher's day-to-day records, the pupil file and the report card. The pupil file is formative in purpose; it is used by teachers to record information on all aspects of the child's learning and development and provides a concise educational history of the child's ongoing progress. In Luxembourg, teachers are required to prepare formative reports (*bilans intermédiaries*) at the end of each trimester. These reports are descriptive in nature, do not contain test scores and are designed to maintain student motivation and

facilitate parents' understanding of student progress towards end-of-cycle objectives (Shewbridge et al., 2012).

Another long-cycle type of formative feedback is the organisation of regular development talks between school leaders/teachers and students and their guardians. In Norway, for example, teachers are expected to maintain documentation of their formative assessment of students and to meet with each student and his/her parents for a discussion of the student's progress once each term. In Sweden also, the IDPs with individualised goals are developed collaboratively in regular "development talks" between the teacher, individual students and their parents.

Such medium- and long-term formative uses of results are important for identifying areas for further improvement, developing broad teaching strategies to address needs identified within the student cohort, planning, allocation of resources, and so on. It can also feed into the school-wide coordination of pedagogical support and remediation for students facing learning difficulties. For example, Blondin and Giot (2011) describe a range of remediation strategies that were put in place in the French Community of Belgium to respond to student learning difficulties identified through formative assessment (Box 4.13).

Box 4.13 Organising remedial education and student orientation based on formative assessment results

In the **French Community of Belgium**, the organisation of formative assessment in all schools is mandated by decree. The importance of training teachers to be able to guarantee early identification of learning difficulties and immediate remediation is emphasised in the Community's Policy Declaration 2009-2014. While remediation strategies vary across schools, there are indications that schools are bringing together individual formative assessment results for the school-wide organisation of remediation activities. Several schools have established "needs-based groups" working on the basis of formative assessment results. Other schools have designated a teacher or member of school leadership to review requests for support, group them according to identified needs and ensure communication between classroom teachers and remedial teachers to provide coherent support to individual students. The decree also mentions the obligation for schools to organise student orientation by bringing together teacher teams, the psychological and social support centres (centres *psycho-médico-sociaux, CPMS*), parents and students. To this end, each school is serviced by a CPMS, whose multi-disciplinary teams offer free and confidential counselling. While the CPMS do not directly participate in the formative assessment of students, they do have the role of providing specific insights regarding adequate support and orientation strategies for students facing difficulties. They ensure coordination between pedagogical support in the classroom and broader psychological, medical, social and professional support for individual students.

Source: Blondin and Giot (2011).

While medium- and long-term strategies are important to ensure consistency of support throughout a student's learning trajectory, research indicates that short-cycle formative assessment – the daily interactions between and among students and teachers – has the most direct and measurable impact on student achievement (Looney, 2011a). In short-cycle interactions, formative assessment is part of the classroom culture, and is seen as an integrated part of the teaching and learning process. Teachers systematically incorporate such formative assessment methods in their course planning – for example, in how they intend to develop classroom discussions and design activities to reveal student

knowledge and understanding. These interactions encompass effective questioning to uncover student misconceptions and identify patterns in student responses, feedback on student performance and guidance on how to close learning gaps, and student engagement in self- and peer-assessment (OECD, 2005a; Wiliam, 2006; Looney, 2011a).

It is difficult to assess in how far teachers across countries are using this type of feedback in their regular classroom interactions with students. In several countries participating in the OECD Review, including those that already have firm foundations for formative assessments, there were concerns about the adequacy of regular teacher feedback. In several countries, despite some policy attention to formative assessment, feedback was conceived by teachers in a narrow way. It was frequently not immediate and tended to be quite limited in the form of marks or brief comments, focused on test results or performance rather than on learning. To be effective, formative assessment results should be used for feedback that is continuous and embodied in the process of teaching and learning itself (Wiggins, 1998).

Diagnostic use of results

Diagnostic assessment is a common form of formative assessment, which often happens as a kind of pre-assessment at the beginning of a study unit. Diagnostic assessment typically focuses on very specific areas of learning and produces fine-grained information about individual student strengths, weaknesses and learning needs. Many diagnostic tools are designed specifically to uncover the causes of students' learning difficulties. The results of diagnostic assessment are typically used to inform future programme planning, design differentiated instruction and deliver remedial programmes for at-risk students. The distinctive feature of diagnostic assessment, *vis-à-vis* formative assessment more generally, is its greater focus on the use of results for individualised intervention and/or remediation.

Pointers for future policy development

This chapter reviewed country approaches to the assessment of individual students in light of available research and evidence. Based on the analysis developed in this chapter, this section provides a range of policy options, which – collectively – have the potential to enhance student assessment frameworks across OECD countries. These pointers for policy development are drawn from the experiences reported in the Country Background Reports, the OECD Country Reviews and the available research literature.

It should be stressed that there is no single model or global best practice of student assessment. The development of practices always needs to take into account country-specific traditions and features of the respective education systems. Not all policy implications are equally relevant for different countries. In a number of cases many or most of the policy suggestions are already in place, while for others they might not apply owing to different social, economic and educational structures and traditions. Different contexts will give rise to different priorities in further developing student assessment policies. In general, there is a need for further research into the impact of different policy approaches to student assessment. The existing evidence base is dominated by research in a few systems with long-established policies on student assessment. As more systems adopt and implement different student assessment policies, there will be a need to collect evidence on how these impact student learning and educational experiences.

Governance

The main conclusion in relation to governance is the need to develop a *coherent* framework for student assessment. Coherence implies that *(i)* the assessment framework is based on well-conceptualised reference documents (curricula, standards and learning progressions); *(ii)* the purposes of different assessment approaches are clearly set and complement each other; and *(iii)* the responsibilities for governing and implementing the assessment framework are well defined.

Establish a coherent framework for student assessment

Across countries there is increasing emphasis on designing and governing coherent assessment frameworks that integrate different types of assessments and use a range of information to make dependable judgements about student learning. Well-designed assessment frameworks can play a key role in building consensus about education goals, standards and criteria to judge proficiency. They can also be a lever to drive innovation in education by signalling the types of learning that are valued. Establishing clarity about the purposes and appropriate uses of different assessments is important to ensure that assessment frameworks optimally contribute to improvements at the classroom, school and system level. Building the assessment competencies of students, teachers and other stakeholders in the education system is crucial to ensuring the effective implementation of such frameworks.

A key governance challenge for countries is to develop a clear vision and strategy for assessment where different approaches developed nationally and locally each serve a clearly defined purpose and the format of the assessment is aligned to that particular purpose. For assessment to be meaningful, it must be well-aligned to the type of learning that is valued. For example, while simple knowledge tests are well-suited to assess the outcomes of traditional teaching approaches based on rote learning and knowledge transfer, such tests are less adequate when it comes to assessing complex competencies. Coherent assessment frameworks should aim to align curriculum, teaching and assessment around key learning goals and include a range of different assessment approaches and formats, along with opportunities for capacity building at all levels.

Develop clear goals and learning progressions to guide student assessment

In all student assessment systems, there is a need for clear external reference points in terms of expected levels of student performance at different levels of education. While it is important to leave sufficient room for teachers' professional judgements in the classroom, it is necessary to provide clear and visible guidance concerning valued learning outcomes in the central curriculum and standards. This is especially important as many curricula now highlight the need for students to acquire complex competencies. The challenge is that such competency goals are often stated in a general way with little guidance for teachers on how they can adapt their teaching and assessment to reach such goals. Hence, clear and concrete illustrations of the type of learning that should be achieved can provide important support.

Teachers can also benefit from specific learning progressions describing the way in which students typically move through learning in each subject area. Learning progressions can provide a picture from beginning learning to expertise and help provide teachers, parents and other stakeholders with concrete images of what to expect in student learning with direct links to the curriculum. Such learning progressions can provide a clear conceptual basis for a consistent assessment framework, along with assessment

tools that are aligned to different stages in the progressions. Clear descriptions and exemplars of expected learning, along with criteria to assess performance can provide further support.

Ensure a good balance between formative and summative assessment

A large body of research highlights the important role of formative assessment in improving learning outcomes. While most OECD countries have now developed policy frameworks to support and promote formative assessment, little information is available regarding the effective and systematic implementation across schools. There is a risk that pressures for summative scores may undermine effective formative assessment practices in the classroom. In fact, assessment systems that are useful for formative purposes are at risk of losing their credibility if high stakes are attached to them. Such tensions between formative and summative assessment need to be recognised and addressed.

Both formative and summative assessments should be well embedded within broader assessment frameworks. While summative assessment and reporting are important at key stages of the education process, it is the daily engagements of teachers' and students' with assessment information that will lead to sustainable improvements in learning outcomes. To support such practice, a national commitment to formative assessment on paper needs to be matched with a strategic approach to professional learning in assessment and concrete support for teachers and schools.

Establish safeguards against an overreliance on standardised assessments

A clear priority in assessment frameworks is the development of reliable measures of student learning outcomes. This effort has now started in most OECD countries with the development of standardised assessments in the main subjects at key stages of education. Standardised central assessments can help clarify learning expectations for all schools and motivate teachers and students to work towards high standards.

At the same time, there is a risk that the high visibility of standardised assessment, especially where high stakes are attached to them, might lead to distortions in the education process, such as excessive focus on teaching students the specific skills that are assessed, narrowing the curriculum, distributing repeated practice tests, training students to answer specific types of questions, adopting rote-learning styles of instruction, allocating more resources to those subjects that are tested, focussing more on students near the proficiency cut score and potentially even manipulation of results. Because of these potential negative effects, it is important to establish safeguards against excessive emphasis on a particular standardised test and to draw on a range of assessment information to make judgements about learning progress.

Because standardised central assessment is a relatively new phenomenon in many OECD countries, it is important to be clear about its purposes, to develop large-scale assessments over time to be able to accommodate the purposes that are reasonable, point out inappropriate uses and provide guidance for the way in which these assessments can be used as part of a broader assessment framework. Assessment systems require research evidence on the extent to which the interpretations of assessment results are appropriate, meaningful, and useful. The role of the standardised assessments should be clearly fixed and the assessments should be continually developed, reviewed and validated to ensure that they are fit for purpose. Validation is a long-term process of accumulating, interpreting, refining and communicating multiple sources of evidence about appropriate interpretation and use of assessment information. Where new standardised assessments

are introduced, it is important that they are first trialled to enable an evaluation of impacts before full-scale implementation. It is also important for national authorities to clarify the kinds of decisions the assessments can provide evidence for and what decisions require other kinds of information.

Share responsibilities for the governance and implementation of assessment frameworks

Several actors are involved in governing and designing assessment frameworks, including different levels of the educational administration, specialised central assessment agencies, the inspectorates, private assessment developers, educational research centres, school leaders, teachers and students. To ensure the coherence of various assessment approaches, it is important that these different actors engage with each other and investments are made in leadership and collaboration around a joint assessment strategy. Responsibilities for the development and administration of assessments need to be shared between schools and external partners.

A balanced combination of teacher-based and external assessments would be most suitable to leverage both central expertise and teacher ownership and to ensure maximum validity and reliability. Learning outcomes that can be readily assessed in external examinations should be covered this way, whereas more complex competencies should be assessed through continuous teacher-based assessment. There are several ways to centrally support the quality and reliability of teacher-based assessment, for example through the use of scoring guides, negotiated scoring criteria, external benchmarks, training for teachers, multiple judgements and external moderation. It is also important to provide a range of nationally validated assessment tools that teachers can use to assess their students reliably when they see fit.

Procedures

The main conclusion in relation to procedures is the importance of developing a *comprehensive* set of assessment approaches. Comprehensiveness implies that the assessment framework uses a range of assessment instruments, formats and methods so that it captures the key outcomes formulated in national learning goals. The framework should allow teachers to draw on multiple sources of evidence in order to form dependable judgements on student learning. Comprehensiveness also means that assessment approaches are inclusive and able to respond to the various needs of *all* learners.

Draw on a variety of assessment types to obtain a rounded picture of student learning

A comprehensive assessment system should include a range of internal and external approaches and make use of different assessment formats to capture a broad range of learning outcomes for different purposes. It is not appropriate to try and serve multiple purposes with a single assessment. It is important, instead, to develop a comprehensive assessment system that is clear about what the various formats and approaches can achieve and ensures that they are used appropriately for their intended purpose. Providing multiple opportunities and formats for student assessment can increase both the validity and reliability of student assessment.

To ensure a broad approach to student assessment, education systems can provide a range of nationally validated assessment tools for different summative and formative purposes. In addition, teachers also need to build their competencies to develop valid and reliable assessment tools corresponding to specific local needs. Particular attention should be paid to ensuring that the breadth of curriculum and learning goals is maintained in student assessment by ensuring that all subject areas and objectives are given certain forms of attention.

Support effective formative assessment processes

While the importance of formative assessment is widely recognised across OECD countries, in many settings there is room for improving teachers' approaches to formative assessment. For example, formative assessment is sometimes understood by teachers as having many small practice tests in view of preparing a final summative assessment, or as providing feedback in the form of praise or encouragement to make more effort. However, for formative assessment to be effective it needs to be independent of the requirement to rate performance, and for feedback to be helpful for student learning, it needs to provide timely, specific and detailed suggestions on the next steps to enhance further learning.

Education authorities can support formative assessment procedures with a range of tools that may help schools in developing systematic approaches. The use of individual development plans for each student can support medium- and long-cycle formative assessment processes. In addition, specific guidelines, workshops, online tools and professional learning opportunities can support effective formative assessment on a daily basis (more on this under "Capacity").

Clarify and illustrate criteria to judge performance in relation to national goals

To assist teachers in their practical assessment work against learning goals, there is also a need to develop support materials, such as scoring rubrics listing criteria for rating different aspects of performance and exemplars illustrating student performance at different levels of achievement. Clear scoring rubrics can make teachers' assessment transparent and fair and encourage students' metacognitive reflection on their own learning. They can be used to define what constitutes excellent work and enable teachers to clarify assessment criteria and quality definitions.

Such guidance can help teachers make accurate judgements about student performance and progress, which is essential to make decisions about how to adapt teaching to students' needs. Teachers also need to acquire skills to develop their own specific objectives and criteria aligned with national learning goals, and should be encouraged to share and co-construct such assessment criteria with students, so that they understand different levels of quality work.

Ensure the consistency of assessment and marking across schools

While most countries set basic requirements regarding the use of particular marking scales and reporting formats, there tend to be large inconsistencies in marking practices across teachers and schools. Such inconsistency in approaches to marking reduces the value of marks as a tool to summarise and inform learning. It is also unfair to students, especially when marks are used to make decisions about their future educational trajectory. Central specifications regarding summative assessment and marking are important to help a consistent application of assessment criteria across schools.

In addition, moderation processes are key to increase the reliability of teacher-based assessment. Moderation involves strategies for quality assurance of assessment judgements, such as teachers cross-marking each other's assessments within a school or across schools, teachers discussing samples of student work in groups or in collaboration with experienced moderators, or a competent authority or examination board externally checking school-based assessments. The objective is to reduce variations in the ways teachers assess students and set marks in order to achieve fairness in student assessment and reporting.

Moderation practices should be encouraged for different types of assessments at all levels of education. It would be beneficial to develop guidelines and support for appropriate approaches to moderation, both within and across schools. Such guidelines should emphasise the importance of moderation as a process for developing assessment confidence and common understandings of assessment standards among teachers, but also as a mechanism to increase the dependability (validity and reliability) of teacher assessments of student performance.

Promote assessment formats that capture valued key competencies

Across OECD countries, there is increasing emphasis in curricula on the importance for students to acquire key 21st century competencies and education systems need to adapt their assessment approaches so that they promote and capture this broader type of learning. To this end, teachers need to be supported in translating competency goals into concrete lesson plans, teaching units and assessment approaches. Professional learning opportunities where teachers can discuss and collaborate in assessing actual student products can contribute to their understanding of broader assessment practices.

In addition, innovative assessment formats should also be developed centrally to complement teacher-made assessments. Due to concerns about reliability and resources, "performance-based" or "authentic" assessments are often challenging to implement on a large scale and in a standardised way. Alternatively, education systems can opt for developing test banks for teachers, which can provide a range of innovative assessment tools for teachers to draw from when their students are ready. Such test banks provide an excellent opportunity to promote innovative assessment tools that have proven successful elsewhere. They can offer a map of assessment items suitable to assess the key areas and competencies outlined in the curriculum.

Another option is to implement innovative assessments that cover larger parts of the curriculum on a sample basis. Sample-based assessments that are applied to a representative proportion of the student cohort allow the assessment of a broader range of curriculum content at relatively low cost while at the same avoiding distortions deriving from potential "teaching to the test". Such assessments may be organised in cycles, assessing a different curriculum area each year and not assessing all students on the same tasks, thus allowing the assessment of a wider range of content without overburdening individual students. While the purpose of such sample-based assessment typically is to monitor the education system (see Chapter 8), they can still be beneficial for individual teachers and students when they receive their results. The tasks of previous years may also be made available for teachers to use in their formative classroom assessment. Where teachers are centrally employed and trained to correct such sample-based assessments, this can constitute a valuable professional learning experience that will also help them in their classroom teaching and assessment practice.

Build on innovative approaches developed in particular education sectors

In many countries, there are some education sectors which have a longer tradition than others in using innovative assessment approaches. Often, there is a stronger tradition in the vocational education and training (VET) sector than in general education programmes to include innovative assessment approaches that may take place in practical and authentic work situations and are connected to real-life challenges that graduates may encounter in the workplace.

Sometimes there is also greater attention to assessing authentic performances in special needs education, second chance education programmes or special programmes for migrant students. In designing assessment approaches for general education programmes, it would be important to pay close attention to innovative assessments developed in other programmes and learn from approaches that have been shown successful and could be integrated and/or adapted. Policy makers should promote communication and collaboration regarding the assessment of competencies across education sectors and programmes, so that mutual learning can be facilitated.

Tap into the potential of ICT to develop sophisticated assessment instruments

Increasingly sophisticated ICT programmes that score open-ended performances, measure students' reasoning processes, examine how students go about thinking through problems and even provide feedback to students have been developed in some settings, predominantly in the United States. While it has always been possible for teachers or external assessors to perform these functions, ICT offers the possibility for large-scale and more cost-effective assessment of complex skills (Mislevy et al., 2001, in Looney, 2009).

While introducing constructed-response items and open-ended performance tasks in large-scale assessments is quite demanding, technology today makes this possible and more affordable. Increased efficiency would allow systems to administer tests, at different points in the school year, with results to be used formatively (as with curriculum-embedded or on-demand assessments). In addition, computer-based assessments can help increase the timeliness of feedback to teachers and students. While in many countries, central assessment systems provide teachers with results several months after the tests were administrated, the use of ICT-based assessment allows providing feedback to teachers very rapidly. With computer-based tests, it is possible to provide teachers and students with their test results the next day, which can foster the use of the test results for adapting teaching and learning for individual student progress.

In addition, computer-based assessments that adapt test items to the level of student performance on previous items can strengthen the diagnostic dimension of large-scale assessments. Only a few countries are using computer-based adaptive tests but these may provide powerful pedagogical tools for teachers. While a standardised test can only provide a snapshot of student achievement in selected targets and subjects, within a discrete area adaptive tests are able to provide a very accurate diagnosis of student performance. As each student sits a different test including questions that are adapted to his/her ability level, this can allow a more thorough diagnostic feedback.

These kinds of assessments are relatively new, and as of yet, relatively limited in number across OECD countries. However, as systems invest more in research to develop appropriate measurement technologies that are able to score complex performances and that reflect models of learning in different domains, development is likely to accelerate.

Ensure that student assessment is inclusive and responsive to different learner needs

Assessment systems should underline the importance of responding to individual learner needs and school community contexts, and design assessment strategies that suit the needs of different learner groups. The objective is to develop an inclusive student assessment system based on the principle that all students have the opportunity to participate in educational activities, including assessment activities, and to demonstrate their knowledge, skills and competencies in a fair way. Hence, teacher assessment practices as well as the content and format of standardised assessments should be sensitive to particular groups of students and avoid biases by socio-economic background, immigrant or minority status, and gender.

While innovative and motivating assessment strategies are important for all students, this is particularly the case for vulnerable students or students at risk of dropping out. Several studies indicate that certain formats of assessment may advantage or disadvantage certain student groups. Hence, to ensure fairness in assessment, it is important to offer a range of different assessment formats and tasks (e.g. test-based, performance-tasks, oral, written).

Dimensions of inclusive assessment, such as the sensitivity to cultural and linguistic aspects of assessment, should also be further included and developed in both initial education and professional development for teachers. The accessibility and lack of bias of standardised assessments for certain groups at risk of underachievement should receive due attention. This requires studies on differential test functioning for particular groups and the provision of specific test accommodations where necessary. It is suggested that quality assurance guidelines are prepared and practices adopted to ensure that assessments are reviewed for their potential bias in these respects.

Capacity

The main conclusion in relation to capacity relates to the need for assessment frameworks to be *participatory*. Student assessment involves a broad range of actors including students, teachers, school leaders, parents, education authorities and assessment agencies. All of these actors need to develop their competencies to ensure that stated objectives are reached, starting with the students themselves.

Put the learner at the centre and build students' capacity to engage in their own assessment

For assessment systems to enhance learning – and not just measure it – students need to be at the centre of the assessment framework. To become lifelong learners, they need to be able to assess their own progress, make adjustments to their understandings and take control of their own learning. Assessment can only lead to improvement in learning outcomes if students themselves take action and use assessment information to close gaps in their own learning. Recent educational research emphasises the importance of assessment as a process of metacognition, where learners become aware of their own thought processes, personally monitor what they are learning and make adaptations in their learning to achieve deeper understanding.

Self-and peer-assessment are powerful processes where students identify standards and criteria to make judgements about their own and their peers' work, which can promote a greater sense of agency and responsibility for their (life-long) learning. But

developing skills for self-assessment and self-regulation takes time and requires structured support by teachers in the classroom. Teachers can use classroom assessment to provide opportunities for students to engage in reflection and critical analysis of their own learning, for example by guiding students in setting learning goals and monitoring their progress towards them; working with them to develop criteria to judge progress; using exemplars and models of good practice and questioning of their own thinking and learning processes. Policy makers can support such practices by developing requirements, guidelines and support regarding learner-centred teaching and assessment.

Maintain the centrality of teacher-based assessment and promote teacher professionalism

Across many countries, there is recognition that teacher professionalism needs to be at the heart of effective assessment for learning. Students will develop their own assessment capacity only if teachers themselves have such capacity and are adequately resourced (Absolum et al., 2009). Placing a strong emphasis on teacher-based assessment has many advantages: it allows for competencies to be measured that are difficult to capture in a standardised assessment, it is embedded in regular coursework and more authentic than a test-based examination and it has greater potential to be used for subsequent improvements in teaching and learning.

However, in order to reach the full potential of teacher-based assessment, it is important for policy makers and stakeholders to adopt a strategic approach to teacher learning in assessment and invest in professional learning opportunities. Assessment capacity should be reflected in teacher standards and be addressed in a coherent way across teacher preparation programmes and publicly funded professional development programmes.

Teachers' assessment capacity can further be built and strengthened through systematic arrangements for moderation of assessments. There is considerable evidence that involving teachers in moderation is a powerful process not only for enhancing consistency but also for enabling teachers to deeply understand student learning objectives and to develop stronger curriculum and instruction. Moderated assessment and scoring processes are strong professional learning experiences that can drive improvements in teaching, as teachers become more skilled at various assessment practices and the use of assessment information to make adjustments to teaching and learning approaches.

Identify assessment priority areas for teacher initial education and professional development

There are variations across countries regarding the areas where teachers need most support to develop effective assessment practice. It is important for policy makers, together with teacher education institutions and stakeholders, to identify the topics related to assessment that are most in need of development within teacher education. Experience from the OECD Review indicates that the following areas require particular attention in many countries.

First, to be able to assess students' progress in developing complex competencies, it is important that teachers learn to develop a variety of assessment approaches and understand different aspects of validity, including what different assessments can and cannot reveal about student learning. Second, for summative teacher-based assessment to be reliable, it is important to provide focussed training on how to make summative

judgements on student performance in relation to national curriculum goals or standards and how to apply marking criteria to very different types of evidence of student learning. Third, for formative assessment to be effective, it is essential that teachers are offered in-depth professional learning opportunities in particular on embedding formative assessment in regular daily teaching practice, co-developing clear criteria for assessment with learners, giving specific, timely and detailed feedback, and creating conditions for students to develop self-monitoring skills. Fourth, to increase equity and fairness in assessment, training should also focus on ensuring that teachers are sensitive to cultural and linguistic aspects of learning and assessment.

Use teacher appraisal and school evaluation processes to help teachers develop their assessment capacity

Teacher appraisal and school evaluation processes can also contribute to identifying those areas of student assessment where teachers most need to develop their skills. Effective models of teacher appraisal and school evaluation specifically focused on teachers' and schools' assessment approaches have been piloted in some settings but are only in the early stages of development. Inspection visits, for example, may contribute to fostering innovation in assessment, by focussing on a wide range of quality indicators (beyond test results) including the capacity of teachers and schools to promote and assess key competencies.

Reporting and use of results

The main conclusion regarding the reporting and use of assessment results is the need for the assessment framework to be *informative*. It needs to produce high-quality information that can be shared with students, parents, school leaders and others with an interest in student learning outcomes. Reporting of assessment information needs to be clear, contextualised, and useful to foster learning and feed into decision making at different levels of the education system.

Develop clear reporting guidelines

Effective reporting is essential to communicate summary statements of achievement to students and their parents, as well as to other teachers within the school. Such records can support co-operation with parents, ensure consistency of support after student transitions to higher levels of education and provide a basis to make decisions about a student's further educational career. However, where there is a lack of transparency and consistency in the ways in which marks and report cards are constructed, the effect of such reporting will be counterproductive for student motivation and future learning.

Clear central reporting guidelines can help build a common understanding around the meaning of marks and the criteria used to establish student performance. They can also help to clarify that information about student behaviour, effort and motivation should not be mixed into performance marks. Keeping these elements separate in reporting allows communicating more accurate information about the student as a complex learner and can provide better indications about how particular learning needs can best be addressed.

Engage parents in education through adequate reporting and communication

Good reporting and communication strategies are important for involving parents in supporting their children's learning and in focussing resources, both at school and at home, on essential learning targets. Hence, reporting needs to be clear and easy to understand,

especially in primary education when parents and teachers can have the greatest impact on a child's learning. While some countries have standardised reporting approaches, others leave it to the local and school level to determine the format of reporting.

To ensure minimum quality requirements, countries could consider providing a template for reporting student achievement and provide guidance materials that teachers can use to report student performance in relation to student learning objectives. Useful information, beyond simple marks, would include details about student progress, strengths, areas of concern, identified needs, recommendations for further learning and illustrative examples.

Ensure transparency and fairness when using assessment results for high-stakes decisions

The results of summative assessment may be used by a range of stakeholders for different purposes and some of these users may have little knowledge of the intended uses of the assessments, the content of the assessments and the evidence concerning the validity of inferences from the assessments. Hence, there is a risk of misuse of assessment results. Also, several reviews of research have found the high stakes use of a single assessment to be strongly related with teachers focusing on the content of the assessments rather than underlying learning goals, administering repeated practice tests, training students for answering specific types of questions and students adopting surface learning techniques. To avoid such negative impacts on teaching and learning and reduce the risk of misuse of results, a number of principles on appropriate use of test results should be respected. For example, students should have multiple opportunities to show their learning, results from a single assessment alone should not be used to make high-stakes decisions about their future learning pathways, and alternative assessments should be considered to replace high-stakes testing, especially for students where there is a high probability that tests give a misleadingly low indication of their competencies.

Promote the regular use of assessment results for improvement

Assessment is closely intertwined with teaching and learning. Whether internal or external, assessment cannot be separated from a vision about the kind of learning that is valued and the teaching and learning strategies that can help students to get there. In turn, there is strong research evidence on the power of assessment to feed forward into new teaching and learning strategies and the strong relationship between assessment for learning and improvement of learning outcomes. To optimise the potential of assessment to improve what is at the heart of education – student learning – policy makers should promote the regular use of assessment results for improvements in the classroom. All types of assessment should have educational value, and be meaningful to those who participate in the assessment. To this end, it is important that all those involved in assessment at the central, local and school level have a broad vision of assessment and of the need to bring together results from different types of assessment activities to form rounded judgements about student learning and the use of assessment information for further improvement.

Notes

1. The use of assessment information to evaluate teachers, schools, school leaders and education systems will be discussed in Chapters 5 (teacher appraisal), 6 (school evaluation), 7 (school leader appraisal) and 8 (education system evaluation) of this report. Hence, in this chapter, standardised assessment is only included to the extent that it is used to make judgements about the learning of individual students – its school evaluation function is discussed in Chapter 6 and its national monitoring function is covered in Chapter 8. International student surveys will be discussed in Chapter 8.

2. See Tables 4.A1.4, 4.A2.4a and 4.A2.4b, as well as Tables 4.A1.5, 4.A2.5a and 4.A2.5b in Annexes 4.A1 and 4.A2.

3. This refers to the state level in Australia, the Community level in the Flemish and French Communities of Belgium, the province/territory level in Canada and the regional government level in Northern Ireland.

4. In Italy, Luxembourg and Portugal, they apply to public schools only, in Austria they apply to public and government-dependent private schools only and in Denmark exception is granted to a few private schools.

5. Except in Italy, Ireland, Luxembourg and New Zealand, where they are not compulsory for all private schools.

6. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision-making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

7. The SAM-scale is available at the website *www.competento.be*. The website provides a non-exhaustive overview and links to screening tools for both education and non-education objectives (e.g. self-tests for students to evaluate their learning style). The website is supported by the Flemish Agency for Entrepreneurial Training (*Vlaams Agentschap voor Ondernemersvorming*).

8. The Educational Quality and Accountability Office of Ontario (EQAO), for example, has chosen option 2 and reports assessment results separately for French and English students. As results are not on the same scale, standard setting and reporting of results must be done separately. More information on EQAO testing in both English and French can be found on the EQAO website: *www.eqao.com/Parents/FAQ.aspx?Lang=E&gr=036*.

9. For instance, there could be a core of items or activities that would be the same, except for translation and adaptation. These items or activities would serve to equate results on other parts of the instrument which have not been translated and which are unique to each language group.

10. In Austria, Canada, Denmark (ISCED 1), Hungary (ISCED 2 and 3), Ireland and Mexico (ISCED 3), they are not compulsory for private schools.

References

Abedi, J. (2004), "The No Child Left Behind Act and English language learners: Assessment and accountability issues", *Educational Researcher*, Vol. 33, No. 1, pp. 4-14.

Abedi, J. and P. Gandara, (2006), "Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy", *Educational Measurement: Issues and Practices*, Vol. 25, No. 4, pp. 36-46.

Abedi, J., C. Hofstetter and C. Lord (2004), "Assessment accommodations for English language learners: Implications for policy-based empirical research", *Review of Educational Research*, Vol. 74, No. 1, pp. 1-28.

Absolum, M., L. Flockton, J. Hattie, R. Hipkins and I. Reid (2009), *Directions for Assessment in New Zealand (DANZ) Report: Developing Students' Assessment Capabilities*, Ministry of Education, Wellington.

ACACA (Australasian Curriculum, Assessment and Certification Authorities) (1995), *Guidelines for Assessment Quality and Equity*, ACACA, *http://acaca.bos.nsw.edu.au/files/pdf/guidelines.pdf*.

AERA, APA, NCME (American Educational Research Association, American Psychological Association, National Council for Measurement in Education) (1999), *Standards for Educational and Psychological Testing*, AERA, Washington D.C.

Alderson, C. and D. Wall (1993), "Does washback exist?", *Applied Linguistics*, Vol. 14, pp. 115-129.

Anderson, B., D.S. MacDonald and C. Sinnemann (2004), "Can measurement of results help improve the performance of schools?", *The Phi Delta Kappan*, Vol. 85, pp. 735-739.

Baartman, L.K.J., T.J. Bastiaens, P.A. Kirschner, C.P.M. Van der Vleuten (2006), "The wheel of competency assessment: Presenting quality criteria for competency assessment programs", *Studies in Educational Evaluation*, Vol. 32, pp. 153-170.

Baker, E.L. and R.L. Linn (2004), "Validity issues for accountability systems", in S.H. Fuhrman and R.F. Elmore (eds.), *Redesigning Accountability Systems for Education*, Teachers College Press, New York, pp. 47–72.

Barneveld, C. van (2008), "Using data to improve student achievement", *What Works? Research into Practice*, Research Monograph No. 15, The Literacy and Numeracy Secretariat, Ontario, *www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/Using_Data.pdf*.

Becker, W. and S. Rosen (1992), "The learning effect of assessment and evaluation in high school", *Economics of Education Review*, Vol. 11, No. 2, pp.107-118.

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

Bennett, R.E. (2001), "How the Internet will help large-scale assessment reinvent itself", *Education Policy Analysis*, Vol. 9, No. 5.

Biggs, J. (1999), *Teaching for Quality Learning at University*, SRHE and Open University Press, Buckingham, UK.

Biggs, J. (1998), "Assessment and classroom learning: A role for summative assessment?", *Assessment in Education: Principles, Policy & Practice*, Vol. 5, No. 1, pp. 103-110.

Biggs, J. (1996), "Enhancing teaching through constructive alignment", *Higher Education*, Vol. 32, pp. 347-364.

Binkley, M., O. Erstad, J. Herman, S. Raizen, M. Ripley with M. Rumble (2010), *Draft White Paper 1: Defining 21st Century Skills*, *http://atc21s.org/wp-content/uploads/2011/11/1-Defining-21st-Century-Skills.pdf*.

Bishop, J.H. (2006), "Drinking from the fountain of knowledge: Student incentive to study and learn – externalities, information problems and peer pressure", *Handbook of the Economics of Education*, Vol. 2, pp. 910-944.

Bishop, J.H. (1999), "Are national exit examinations important for educational efficiency?", S*wedish Economic Policy Review,* Vol. 6, No. 2, pp. 349-398.

Bishop, J.H. (1997), "The effect of national standards and curriculum-based examinations on achievement, *American Economic Review,* Vol. 87, No. 2, pp. 260-264.

Black, P. (1993), "Formative and summative assessments by teachers", *Studies in Science Education*, Vol. 21, pp. 49-97.

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

Blondin, C. and B. Giot (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the French Community of Belgium,* Université de Liège, *www.oecd.org/edu/evaluationpolicy*.

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

Bloom, B. (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*, by a committee of college and university examiners.

Boud, D. (1986), *Implementing Student Self Assessment*, Higher Education Research and Development Society of Australia (HERDSA), Sydney.

Boulet, M., G. Simard and D. Demelo (1990), "Formative evaluation effects on learning music", *Journal of Educational Research*, Vol. 84, pp. 119-125.

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

Butler, R. (1988), "Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance", *British Journal of Educational Psychology*, Vol. 58, pp. 1-14.

Butler, D. and P.H. Winne (1995*), "*Feedback as self-regulated learning: A theoretical synthesis"*, Review of Educational Research*, Vol. 65, pp. 245-281.

Camara, W. (n.d), "Standards for Educational and Psychological Testing", The Gale Group, *www.education.com/partner/articles/galegrp/*

Chung, G.K.W.K. and E.L. Baker (2003), "Issues in the reliability and validity of automated scoring of constructed responses" in M.D. Shermis and J. Burstein (eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Lawrence Erlbaum, Mahwah, New Jersey, pp. 23-40.

Chung, G.K.W.K. et al. (2001), *Knowledge Mapper Authoring System Prototype*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Cizek, G. J. (1997), "Learning, achievement, and assessment: Constructs at a crossroads", in G. D. Phye (ed.), *Handbook of Classroom Assessment: Learning, Achievement, and Adjustment* (p 1-32), Academic Press, San Diego.

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

Clarke, M., W. Haney and G. Madaus (2000), "High stakes testing and high school completion", *National Board on Educational Testing and Public Policy Statement*, Vol. 1, No. 3, School of Education, Boston College, *www.bc.edu/research/nbetpp/publications/v1n3.html*.

Clarke, M., A. Shore, K. Rhoades, L. Abrams, J. Miao and J. Li (2003), *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States,* Lynch School of Education, Boston College, Massachusetts, National Board on Educational Testing and Public Policy.

Crooks, T.J. (2004), "Tensions between assessment for learning and assessment for qualifications", paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies (ACEAB), Fiji, March 2004.

Crooks, T.J. (1988), "The impact of classroom evaluation practices on students", *Review of Educational Research*, Vol. 58, pp. 438-481.

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

Darling-Hammond, L. (2010), *Draft White Paper 5: Policy Frameworks for New Assessments,* Assessment and Teaching of 21st Century Skills (ATCS), *http://atc21s.org/wp-content/uploads/2011/11/5-Policy-Framework.pdf.*

Darling-Hammond, L. (1998), "Policy and change: Getting beyond bureaucracy" in A. Hargreaves, A. Lieberman, M. Fullan and D. Hopkins (eds.), *International Handbook of Educational Change*, Kluwer Academic Publishers, Dordrecht.

Dee, T.S. and B.A. Jacob (2006), "Do high school exit exams influence educational attainment or labor market performance?", *NBER Working Paper 12199*, National Bureau of Economic Research, Cambridge, Massachusetts, *www.nber.org/papers/w12199*.

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

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

Earl, L. (2003), "Assessment as learning: Using classroom assessment to maximize student learning", in T.R. Guskey and R.J. Marzano (eds.), *Experts in Assessment Series*, Corwin Press, Thousand Oaks, California.

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

Earl, L. and S. Katz (2008), "Getting to the core of learning: Using assessment for self-monitoring and self-regulation", in S. Swaffield (2008), *Unlocking Assessment: Understanding for Reflection and Application*, Routledge, Taylor and Francis, London.

Elmore, R. (1996), "Getting to scale with good educational practice", *Harvard Educational Review*, Vol. 66, No. 1.

Elmore, R. and G. Sykes (1992), "Curriculum policy", in P.W. Jackson (ed.), *Handbook of Research on Curriculum*, Macmillan, New York.

Elshout-Mohr, M., R. Oostdam and M. Overmaat (2002), "Student assessment within the context of constructivist educational settings", *Studies in Educational Evaluation*, Vol. 28, No. 4, pp. 369-390.

EPPI (2005), *Research Evidence on the Impact on Students of Self- and Peer-assessment*, Review conducted by the Assessment Review Group, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI), London.

EPPI (2004), *A Systematic Review of the Evidence of the Reliability and Validity of Assessment by Teachers for Summative Purposes*, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.

EPPI (2002), *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.

Eraut, M. (1990), "Identifying the knowledge which underpins performance" in H. Black and A. Wolf, *Knowledge and Competence: Current Issues in Education and Training*, Careers and Occupational Information Centre, Employment Department, London, UK.

European Commission (2012), *Assessment of Key Competences in Initial Education and Training: Policy Guidance*, Commission Staff Working Document, *http://ec.europa.eu/education/news/rethinking/sw371_en.pdf*.

European Commission (2011a), *Evidence on the Use of ICT for the Assessment of Key Competences*, European Commission, Brussels.

European Commission (2011b), *Assessment of Key Competences: Policy Handbook*, European Commission, Brussels.

Eurydice (2009a), *Key Data on Education in Europe 2009*, European Commission, Brussels.

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

Ewell, P.T. (2005), *Applying Learning Outcomes Concepts to Higher Education: An Overview*, prepared for the University Grants Committee.

Figlio, D.N. and M.E. Lucas (2004), "Do high grading standards affect student performance?", *Journal of Public Economics*, Vol. 88, pp. 1815-1834.

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

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

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

Frederiksen, J. and B. White (2004), "Designing assessment for instruction and accountability: An application of validity theory to assessing scientific inquiry", in M. Wilson (ed.), *Towards Coherence between Classroom Assessment and Accountability*, 103rd Yearbook of the National Society for the Study of Education, Part II, Chicago, Illinois.

Gagné, R.M. (1977), *The Conditions of Learning*, Holt, Rinehart & Winston, Inc., New York.

Gilmore, A. (2008), *Assessment Review Paper 8: Professional Learning in Assessment*, Ministry of Education, Wellington.

Gipps, C. and G. Stobart (2009), "Fairness in assessment", in C. Wyatt-Smith and J. Cumming (eds.), *Educational Assessment in the 21st Century*, Springer.

Gipps, C. and G. Stobart (2004), "Fairness in assessment", in *Perspectives on Pupil Assessment*, General Teaching Council for England (GTC), London.

Graham, A.E. and T.A. Husted (1993), "Understanding state variations in SAT scores", *Economics of Education Review,* Vol. 12, No. 3, pp. 197-202.

Greaney, V. and T. Kellaghan (1995), *Equity Issues in Public Examinations in Developing Countries,* World Bank, Washington DC.

Hajdúková, V., M. Šnídlová, A. Vykydalová, T. Lukács, A. Mentel, M. Pokorný, Z. Juščáková and R. Kanovská (forthcoming), *OECD Review on Evaluation* and *Assessment Frameworks for Improving School Outcomes: Country Background Report for the Slovak Republic,* National Institute for Certified Educational Measurements (NÚCEM), Bratislava, *www.oecd.org/edu/evaluationpolicy*.

Hamilton, L.S., B.M. Stecher, J. Marsh, J.S. McCombs, A. Robyn, J. Russell, S. Naftel and H. Barney (2007), *Implementing Standards-based Accountability Under No Child Left Behind: Responses of Superintendents, Principals, and Teachers in Three States,* RAND, Santa Monica, California.

Hargreaves, D.J., M.J. Galton and S. Robinson (1996), "Teachers' assessments of primary children's classroom work in the creative arts", *Educational Research*, Vol. 38, pp. 199-211.

Harlen, W. (2007), "Criteria for evaluating systems of student assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 15-28.

Harlen, W. (2005), "Teachers' summative practices and assessment for learning – tensions and synergies", *The Curriculum Journal*, Vol. 16, No. 2, pp. 207-223.

Harlen, W. (2004), "Can assessment by teachers be a dependable option for summative purposes?", *Perspectives on Pupil Assessment*, General Teaching Council for England (GTC), London.

Harlen, W. and M. James (1997), "Assessment and learning: Differences and relationships between summative and formative assessment", *Assessment in Education: Principles, Policy and Practice*, Vol. 4, No. 3, pp. 365-379.

Hattie, J. and H. Timperley (2007), "The power of feedback", *Review of Educational Research*, Vol. 77, No. 1, pp. 81-112.

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

Herl, H.E. et al. (1999), "Reliability and validity of a computer-based knowledge mapping system to measure content understanding", *Computers in Human Behavior*, Vol. 15, pp. 315-334.

Herman, J.L., E. Osmundson and D. Silver (2010), *Capturing Quality in Formative Assessment Practice: Measurement Challenges*, CRESST Report 770, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Hidi, S. and J.M. Harackiewicz (2000), "Motivating the academically unmotivated: A critical issue for the 21st century", *Review of Educational Research*, Vol. 70, No. 2, pp. 151-179.

Hopper, M. (2001), *The Implications of Accommodations in Testing Students with Disabilities*, Educational Resources Information Center, *www.eric.ed.gov/ERICWebPortal/detail?accno=ED463627*.

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

Irish Department of Education and Skills (2011*), Literacy and Numeracy for Learning and Life: The National Strategy to Improve Literacy and Numeracy among Children and Young People 2011-2020*, Dublin, *www.education.ie.*

Istance, D. and H. Dumont (2010), "Future directions for learning environments in the 21st century" in Dumont, H., D. Istance and F. Benavides (eds.), *The Nature of Learning*, OECD Publishing, Paris.

Jacob, B.A. (2005). "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools", *Journal of Public Economics*, Vol. 89, No. 5, pp. 761-796.

Kärkkäinen, K. (2012), "Bringing about curriculum innovations: Implicit approaches in the OECD area", *OECD Education Working Papers*, No. 82, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Kellaghan T., G. Madaus and A. Raczek (1996), *The Use of External Examinations to Improve Student Motivation,* American Educational Research Association (AERA), Washington DC.

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

Klein, S., L. Hamilton, D. McCaffrey and B. Stencher (2000), *What Do Test Scores in Texas Tell Us?,* Issue paper, Rand Education.

Klemp, G.O. (1977), "Three factors of success" in D.W. Vermilye (ed.), *Relating Work and Education*, Jossey-Bass, San Francisco, California.

Klenowski, V. (2009), "Australian Indigenous students: Addressing equity issues in assessment", *Teaching Education*, Vol. 20, No. 1, March 2009, pp. 77-93.

Koh, K. H., C. Tan, P. T. Ng (2012), "Creating thinking schools through authentic assessment: The case of Singapore", *Educational Assessment, Evaluation and Accountability*, Vol. 24, No. 2, pp 135-149.

Kohn, A. (1994), *The Risks of Rewards,* ERIC Clearinghouse of Elementary and Early Childhood Education, Illinois, US.

Kolb, D.A. (1981), "Learning styles and disciplinary differences", in A.W. Chickering (ed.), *The Modern American College*, Jossey-Bass, San Francisco, California.

Köller, O. (2001), "Mathematical world views and achievement in advanced mathematics: Findings from the TIMSS population III", *Studies in Educational Evaluation*, Vol. 27, pp. 65-78.

Koretz, D., R.L. Linn, S.B. Dunbar and L.A. Shepard (1991), "The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests", paper presented at the annual meeting of the American Educational Research Association (AERA), Chicago, 11 April 1991.

Lewis, A. (1998), *Assessing Student Achievement: Search for Validity and Balance*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Linn, R.L. (2000), "Assessments and accountability", *Educational Researcher*, Vol. 29, No. 2, pp. 4–16.

Linn, R.L., E.L. Baker and S.B. Dunbar (1991), *Complex, Performance-Based Assessment: Expectations and Validation Criteria*, CSE Technical Report 331.

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

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

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

Lucas, B. and G. Claxton (2009), *Wider Skills for Learning: What are they, How can they be Cultivated, How could they be Measured and Why are they Important for Innovation?*, Centre for Real-World Learning.

Luedemann, E. (2011), "Intended and unintended short-run effects of the introduction of central exit exams: Evidence from Germany", in E. Luedemann, *Schooling and the Formation of Cognitive and Non-cognitive Outcomes*, Chapter 4, ifo Beiträge zur Wirtschaftsforschung 39, ifo Institut, Munich.

Madaus, G.F. and L.M. O'Dwyer (1999), "A short history of performance assessment: Lessons learned", *The Phi Delta Kappan*, Vol. 80, No. 9, pp. 688-695.

Marsh, C.J. and G. Willis (2007), *Curriculum: Alternative Approaches, Ongoing Issues*, Pearson Education Inc., New Jersey.

Marzano, R.J. (2001), *Designing a New Taxonomy of Educational Objectives*, Corwin Press.

Masqud, M. and C. Pillai (1991), "Effect of self-scoring on subsequent performances in academic achievement tests", *Educational Research,* Vol. 33, pp. 151–154.

McCurdy, B. and E. Shapiro (1992), "A comparison of teacher monitoring, peer monitoring, and self monitoring with curriculum based measurement in reading among students with learning disabilities", *Journal of Special Education,* Vol. 22, pp. 162–180.

McDonald, B. and D. Boud (2003), "The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations", *Assessment in Education*, Vol. 10, No. 2, pp. 209–220.

MCEETYA (2008), *Melbourne Declaration on Educational Goals for Young Australians*, Ministerial Council on Education, Employment, Training and Youth Affairs, Canberra.

McFarlane, A. (2003), "Editorial. Assessment for the digital age", *Assessment in Education*, Vol. 10, pp. 261–266.

McGehee, J.J. and L.K. Griffith (2001), "Large-scale assessments combined with curriculum alignment: Agents of change", *Theory into Practice*, Vol. 40, pp. 137-144.

McMurrer (2007), *Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era*, Center on Education Policy, *www.cep-dc.org/displayDocument.cfm?DocumentID=312*.

Merrett, J. and F. Merrett (1992), Classroom management for project work: An application of correspondence training, *Educational Studies*, Vol. 18, No. 1, pp. 3-10.

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

Mislevy, R.J., L.S. Steinberg, F.J. Breyer, R.G. Almond and L. Johnson (2001), *Making Sense of Data From Complex Assessments*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

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

Natriello, G. (1987), "The impact of evaluation processes on students", *Educational Psychologist*, No. 22, pp. 155-175.

Nayral de Puybusque, A. (forthcoming), "Equity in student assessment: A literature review", *OECD Education Working Papers*, OECD Publishing, Paris.

Newton, P.E. (2007), "Clarifying the purposes of educational assessment", *Assessment in Education*, Vol. 14, No. 2, pp. 149-170.

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

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

Nusche, D. (2008), "Assessment of learning outcomes in higher education: A comparative review of selected practices", *OECD Education Working Papers*, No. 15, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

Nusche, D. (forthcoming), "Summative assessment: What's in it for students?", *OECD Education Working Papers*, OECD Publishing, Paris, *www.oecd.org/edu/workingpapers*.

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

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

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

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

OECD (2012b), *Grade Expectations: How Marks and Education Policies Shape Students' Ambitions*, OECD, Paris.

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

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

OECD (2010b), *Learning for Jobs*, OECD Reviews of Vocational Education and Training, OECD Publishing, Paris.

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

OECD (2007), *Qualifications Systems: Bridges to Lifelong Learning*, OECD Publishing, Paris.

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

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

Ou, D. (2010), "To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam", *Economics of Education Review*, Vol. 29, pp. 171-186.

Papay, J.P., R.J. Murnane and J.B. Willett (2008), "The consequences of high school exit examinations for struggling low-income urban students: Evidence from Massachusetts", *Working Paper 14186*, National Bureau of Economic Research, Massachusetts, *www.nber.org/papers/w14186.*

Pascarella, E.T. and P.T. Terenzini (2005), *How College Affects Students. A Third Decade of Research*, Vol. 2, Jossey-Bass, San Francisco, California.

Pellegrino, J.W., N. Chudowsky and R. Glaser (2001), *Knowing What Students Know*, National Academy Press, Washington, DC.

Pepper, D. (2011), "Assessing key competences across the curriculum – and Europe", *European Journal of Education*, Vol. 46, No. 3.

Popham, J. (1991), "Why standardized tests don't measure educational quality", *Educational Leadership,* Vol. 56, No. 6, pp. 8-15.

Posner, G.J. (1992), *Analyzing the Curriculum*, McGraw-Hill, US.

Pryor, J. and H. Torrance (1998), "Formative assessment in the classroom: Where psychological theory meets social practice", *Social Psychology of Education*, Vol. 2, No. 2, pp. 151-176.

Ravela, P. (2009), "Consignas, devoluciones y calificaciones: los problemas de la evaluación en las aulas de educación primaria en América Latina", Páginas de Educación, Vol. 2, Universidad Católica del Uruguay, Montevideo, pp. 49-89, *www.ucu.edu.uy/Default.aspx?tabid=1267*.

Rawlins, P., J. Brandon, J. Chapman, L. Leach, G. Neutze, A. Scott and N. Zepke (2005), "Standards-based assessment in the senior secondary school: A review of the literature", *New Zealand Journal of Teachers' Work*, Vol. 2, Issue 2, pp. 107-115.

Reynolds, C.R. (1998), "Cultural bias in testing of intelligence and personality", in C. Belar (ed.), *Comprehensive Clinical Psychology*: *Sociocultural and Individual Differences*, Elsevier Science, New York.

Ripley, M. (2009), "Transformational computer-based testing", in F. Scheuermann and J. Björnsson (eds.), *The Transition to Computer-Based Assessment: New Approaches to Skills Assessment and Implications for Large-scale Testing,* Office for Official Publications of the European Communities, Luxembourg.

Rivera, C., E. Collum, L. Schafer Willner and J.K. Sia Jr. (2006), "Analysis of state assessment policies regarding the accommodation of English language learners", in C. Rivera and E. Collum (eds.), *State Assessment Policy and Practice for English Language Learners,* Lawrence Erlbaum Associates, Mahwah, New Jersey.

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

Ross, J.A., C. Rolheiser and A. Hogaboam-Gray (1999), "Effects of self-evaluation training on narrative writing", *Assessing Writing*, Vol. 6, pp. 107-132.

Rothman, R. (2010), *Principles for a Comprehensive Assessment System*, Policy Brief, Alliance for Excellence in Education, Washington, DC.

Rupp, A.A. and N.K. Lesaux (2006), "Meeting expectations?: An empirical investigation of a standards-based assessment of reading comprehension", *Educational Evaluation and Policy Analysis*, Vol. 28, No. 4, pp. 315-333.

Rychen, D.S. (2004), *Key Competencies for All: An Overarching Conceptual Frame of Reference*, DeSeCo Strategy Paper on Key Competencies, *www.unal.edu.co/diracad/competencias/KeyCompetences.pdf*.

Sadler, R. (1989), "Formative assessment and the design of instructional systems", *Instructional Science*, Vol. 18, pp. 119-144.

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

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

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

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

Sattler, J.M. (1992), *Assessment of Children*, 3rd Edition, Jerome M. Sattler, San Diego.

Sawyer R., S. Graham and K. Harris (1992), "Direct teaching, strategy instruction and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities", *Journal of Educational Psychology,* Vol. 84, pp. 340–350.

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

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

Sebba, J. and G. Maxwell (2005), "Queensland, Australia: An outcomes-based curriculum", in *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD Publishing, Paris, pp. 191-204.

Shavelson, R.J., G.P. Baxter and J. Pine (1990), "What alternative assessments look like in science", paper presented at the Office of Educational Research and Improvement Conference: The Promise and Peril of Alternative Assessment, Washington, October 1990.

Shepard, L. (2006), "Classroom assessment", in R. Brennan (ed.), *Educational Measurement*, ACE/Praeger, Westport, pp. 623-646.

Shepard, L.A. and K. Cutts-Dougherty (1991), "Effects of high-stakes testing on instruction", paper presented at the annual meeting of the Educational Research Association.

Shewbridge, C., M. Ehren, P. Santiago and C. Tamassia (2012), *OECD Reviews of Evaluation and Assessment in Education: Luxembourg,* OECD Publishing, Paris, *www.oecd.org/edu/evaluationpolicy*.

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

Shultz, K.S. and D.J. Whitney (2005), *Measurement Theory in Action*, Sage Publications, Thousand Oaks.

Sims, D. (2008), "Strategic responses to school accountability measures: It's all in the timing", *Economics of Education Review*, Vol. 27, No. 1, pp. 58-68.

Singh, M. (2003), "Understanding life skills", background paper prepared for the Education for All Global Monitoring Report 2003/4 "Gender and Education for All: The Leap to Equality", UNESCO Institute for Education, Hamburg, *http://unesdoc.unesco.org/images/0014/001469/146963e.pdf*.

Sjögren, A. (2009), "The long run consequences of being graded in elementary school", paper presented at the Final Conference of the RTN Network "Economic of Education and Education Policy in Europe" (EEEPE), hosted by the Centre for Economic Performance, LSE, London.

Slomp, D. (2008), "Harming not helping: The impact of a Canadian standardized writing assessment on curriculum and pedagogy", *Assessing Writing*, Vol. 13, No. 3, pp. 180-200.

Somerset, A. (1996), "Examinations and educational quality", in A. Little and A. Wolf, *Assessment in Transition, Learning Monitoring and Selection in International Perspective*, Pergamon, Elsevier Science Ltd., Oxford.

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

Stecher, B. and S. Barron (2001), "Unintended consequences of test-based accountability when testing in 'Milepost' grades", *Educational Assessment,* Vol. 7, No. 4, pp. 259–81.

Stiggins, R. (2005), "From formative assessment to assessment for learning: A path to success in standards-based schools", *Phi Delta Kappan*, Vol. 87 (4), pp. 324-328.

Stiggins, R.J. (1999), "Assessment, student confidence and school success", *Phi Delta Kappan*, Vol. 81, No. 3, pp. 191-198.

Stiggins, R.J. (1987), "Design and development of performance assessments", *Education Measurement: Issues and Practice*, Vol. 6 No. 3, pp. 33-42.

Stiggins, R. and J. Arter (2007), *Classroom Assessment for Student Learning: Doing it Right – Using It Well,* Pearson Education, Toronto.

Stiggins, R.J., M.M. Griswold and K.R. Wikelund (1989), "Measuring thinking skills through classroom assessment", *Journal of Educational Measurement*, Vol. 26, pp. 233-246.

Swaffield, S. (2008), "Feedback: The central process in assessment for learning", in S. Swaffield (ed.) *Unlocking Assessment: Understanding for Reflection and Application*, Routledge, Taylor and Francis Group, London and New York.

Thompson, S.J., A.B. Morse, M. Sharpe and S. Hall (2005), *Accommodations Manual: How to Select, Administer, and Evaluate Use of Accommodations for Instruction and Assessment of Students with Disabilities*, The Council of Chief State School Officers, *www.ccsso.org/Documents/2005/Accommodations_Manual_How_2005.pdf*.

Thurlow, M.L., L.L. Christensen and K.E. Lail (2008), "An analysis of accommodations issues from the standards and assessments peer review", *NCEO Technical Report 51*, National Center on Educational Outcomes, *www.cehd.umn.edu/nceo/OnlinePubs/Tech51/Technical51.pdf*.

Timperley, H. (2011), *Realising the Potential of Professional Learning*, University Press, Berkshire.

Timperley, H., A. Wilson, H. Barrar and I. Fung (2008), *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration (BES),* New Zealand Ministry of Education, Wellington.

van de Vijver, F. and N.K. Tanzer (2004), "Bias and equivalence in cross-cultural assessment: An overview", *Revue Européenne de Psychologie Appliquée*, No. 54, pp. 119-135.

Volante, L. (2008), "Equity in multicultural student assessment", *Journal of Educational Thought*, Vol. 42, No. 1, pp. 11-26.

Volkwein, J.F. (2003), "Implementing outcomes assessment on your campus", *Research and Planning EJournal*, Vol. 1, No.1.

Wandall, J. (2010), *National Tests in Denmark: CAT as a Pedagogical Tool*, Danish National School Agency, Copenhagen.

Wang, J. and M.S. Brown (2007), "Automated essay scoring versus human scoring: A comparative study", *The Journal of Technology, Learning, and Assessment"*, Vol. 6, No. 2, *www.jtla.org*.

Westbury, I. (2007), "Making curricula: Why do states make curricula, and how?" in M. Connelly, M.F. He and J.A. Phillion (eds.), *Handbook of Curriculum and Instruction*, Sage, Thousand Oaks, California.

Whiting, G. and D. Ford (n.d.), *Cultural Bias in Testing*, The Gale Group, *www.education.com/reference/article/cultural-bias-in-testing/*.

Wiggins, G. (1998), *Educative Assessment – Designing Assessments to Inform and Improve Student Performance*, Joey-Bass, San Francisco.

Wiliam, D. (2006), "Formative assessment: Getting the focus right", *Educational Assessment*, Vol. 11, pp. 283-289.

Wiliam, D., C. Lee, C. Harrison and P.J. Black (2004), "Teachers developing assessment for learning: Impact on student achievement", *Assessment in Education*, Vol. 11, pp. 49-65.

Wiliam, D. (2010), "The role of formative assessment in effective learning environments", in Dumont, H., D. Istance and F. Benavides (eds.), *The Nature of Learning*, OECD, Paris.

Woessmann, L., E. Luedemann, G. Schuetz and M.R. West (2009), *School Accountability, Autonomy and Choice around the World*, Edward Elgar, Cheltenham, UK.

Annex 4.A1 Features of student assessment frameworks in lower secondary education

The tables below provide information on features of student assessment frameworks in lower secondary education in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Part of the information was supplied by countries through a questionnaire specifically developed by the OECD Review. The rest of the information is based on data provided by OECD *Education at a Glance 2011*.

Additional material on features of student assessment frameworks in primary and upper secondary education is available on line at *http://dx.doi.org/10.1787/9789264190658-8-en*.

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

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

General notes

Australia: Australia is a federation of eight states and territories. There are differences in policy frameworks for evaluation and assessment across states and territories as well as between public (government) and private (non-government) schools.

Belgium (Fl., Fr.): In Belgium, education policy is the responsibility of each Community. The terms "national" and "central", therefore, refer to the highest educational authorities (Ministries of Education) of the Flemish and French Communities of Belgium.

Belgium (Fl.): For public schools, the school organising bodies are typically the central educational authority (Flemish Community) and provincial/regional and local educational authorities (provinces, cities, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Belgium (Fr.): For public schools, the school organising bodies (education networks) are typically the central educational authority (French Community) and provincial and local educational authorities (provinces, municipalities). For government-dependent private schools, the school organising bodies are private entities such as religious communities or associations.

Canada: Canada comprises ten provinces and three territories. Provincial/territorial education authorities refer to the highest level of educational authorities in Canada, as there is no federal/central department of education. There are differences in policy frameworks for evaluation and assessment across provinces and territories.

Chile: For public schools, the school organising bodies (sustainers) are typically local educational authorities (municipalities).

Czech Republic: For public schools, the school organising bodies are typically local education authorities at ISCED levels 1 and 2 and regional education authorities at ISCED level 3.

Hungary: For public schools, the school organising bodies (maintainers) are typically local and regional educational authorities. After two decades of decentralisation, Hungary is experiencing a trend towards a larger degree of central decision making in education. Following new legislation passed in 2011 and 2012, schools and other public educational institutions, with the exception of those maintained by the private sector and religious authorities, are subject to direct governance by central authorities from 2013 onwards. It should be noted that information about Hungary in this chapter refers to the period prior to this reform.

Ireland: School boards of management comprise members external to the school such as representatives of the patron and of the local community as well as members internal to the school such as the principal, teacher representatives and parent representatives.

Netherlands: In principle, all schools are government funded. Depending on their denomination, they can have a private (religious or pedagogy-based) or public character. For public schools, school organising bodies (competent authorities) can be local educational authorities (municipal authorities), a local governing committee with transferred powers of the municipality, or a public foundation or corporation. School organising bodies for private schools can be groups of parents, foundations or corporations.

New Zealand: School Boards of Trustees typically comprise elected members from the school community, the principal, a staff representative and a student representative (in secondary schools).

Spain: Responsibilities for education are shared between the central educational authority (Ministry of Education, Culture and Sport [*Ministerio de Educación, Cultura y Deporte*]) and state educational authorities (Regional Ministries or Departments of Education of the Autonomous Communities [*Comunidades Autónomas*]). The central educational authority executes the general guidelines of the government on education policy and regulates the basic elements or aspects of the system. The Autonomous Communities develop the central regulations and have executive and administrative competences for managing the education system in their own territory. State educational authorities refer to educational authorities at the highest level of the Autonomous Communities. Throughout the tables, the Autonomous Communities are referred to as "state educational authorities".

United Kingdom (Northern Ireland): Following political agreement to the devolution of certain policy and legislative powers from the United Kingdom government at Westminster to a local Assembly in 1999, legislative responsibility for education in Northern Ireland was one of the functions devolved to the Assembly and to a locally elected Minister for Education. The Department of Education, Northern Ireland, provides the central governance and management of education in Northern Ireland and is referred to as the "central level" throughout the tables.

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012)

This table describes central/state policy frameworks for internal summative assessment at ISCED level 2, i.e. forms of student assessment that are not standardised, but designed and marked by students' own teachers and implemented as part of regular classroom instruction in schools. This table focuses on summative forms of internal assessment, i.e. assessment designed to provide a summary statement of student achievement at a particular point in time in order to record, mark or certify achievements that may count towards a student's year-end or final marks. Summative assessment usually occurs at the end of a learning unit, term, school year, or educational level.

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Australia	General	Yes, at the state level for all schools	State agency	State curriculum goals; state standards[1]	Moderation of marking; availability of state guidance materials for marking student performance on the examination	100%
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)	State agency	National standards	Moderation of marking; availability of state guidance materials for marking student performance on the examination	100%
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	Central education authority or government	National curriculum goals	Competences for marking are included in initial teacher education	100%
Belgium (Fl.)	All programmes	No, but certain basic requirements are set at the central level for all schools[2]	Inspectorate	National curriculum goals (attainment targets and developmental objectives)	None	m
Belgium (Fr.)	All programmes	No, but certain basic requirements are set at the central level for all schools[3]	Inspectorate	National curriculum goals	Availability of guidance materials for marking student performance in the examination (provided either by education authorities or school organising bodies [education networks])	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Canada	All programmes	No, but certain basic requirements are set at the provincial/territorial level for public schools and government-dependent private schools only	Provincial/territorial education authorities or governments and a provincial agency in one province (Ontario)	Varies across provinces/territories	Varies across provinces/territories	0%
Chile	All programmes	Yes, at the central level for all schools[4]	School (school principal and teachers)	National curriculum goals; national standards	None	100%
Czech Republic	All programmes	No, but certain basic requirements are set at the central level for all schools[5]	School board or committee; Czech School Inspectorate (CSI)	National curriculum goals; national standards	None	100%
Denmark	All programmes	Yes, at the central level for all schools[6]	Central agency (National Agency for Quality and Supervision)	Binding national objectives (year-level objectives and end objectives for compulsory education)	Availability of national guidance materials for marking student performance on the examination (performance criteria)	50%
Estonia	General	Yes, at the central level for all schools	School (school principal)	National curriculum goal; national standards	None	100%
	Pre-voc and voc	No	a	a	a	a
Finland	All programmes	Yes, at the central level for all schools[7]	Teachers and school leaders	Final assessment criteria in each subject as part of the national core curriculum[8]	Availability of national performance criteria for the transition points in each subject	Not specified in the policy framework
France	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Hungary	All programmes	Yes, at the central level for all schools	School organising bodies (maintainers) and school leaders	National curriculum standards	None	100%
Iceland	All programmes	No, but certain basic requirements are set at the central level for all schools[9]	School board	National curriculum goals	None	m
Ireland	All programmes	Yes, at the central level for public and government-dependent private schools only	Teachers and school leaders	National curriculum goals	Availability of national guidance material for assessing student performance[10]	100%
Israel	All programmes	No	a	a	a	a
Italy	All programmes	Yes, at the central level for all schools[11]	School board or committee (includes the school principal and all teachers)	National curriculum goals[11]	None	100%
Korea	All programmes	Yes, at the central level for all schools	Respective subject committee within the school	National curriculum goals	Decision made by the School Mark Management Committee within the school	100%
Luxembourg	All programmes	Yes, at the central level for public schools	Central education authorities or government; school principal	National curriculum goals	None	100%
Mexico	All programmes	Yes, at the central level for all schools	State education authorities or governments	National curriculum goals	None	100%
Netherlands	All programmes	Yes, at the central level for all schools	School organising bodies (competent authorities)	National curriculum goals	Moderation of marking (within schools); Inspectorate monitors differences between school-based assessment and central examinations at the end of the cycle	End of year: 100%; End of cycle: 50%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
New Zealand	All programmes	Yes, at the central level for state schools and state-integrated schools only	School Board of Trustees	National curriculum goals	Availability of national guidance materials for marking student performance (performance criteria, exemplars); moderation of marking (within and between schools)	100%
Norway	All programmes	No, but certain basic requirements are set at the central level for all schools[12]	Local education authorities	National curriculum goals	None	100%
Poland	All programmes	Yes, at the central level for all schools[13]	School principal	National standards; curriculum decided at the school level	None	100%
Portugal	General	Yes, at the central level for public schools only	National education authority; pedagogical council	National curriculum goals	Moderation of marking within schools (criteria approved by the pedagogical council)	70%
	Pre-voc and voc	No	a	a	a	a
Slovak Republic	General	No, but certain basic requirements are set at the central level for all schools	School principal	National standards, national education programme	None	m
	Pre-voc and voc	No	a	a	a	a
Slovenia	All programmes	Yes, at the central level for all schools	School principal	National curriculum goals	None	100%

Table 4.A1.1 Internal summative assessment frameworks at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates internal summative assessment?	Who is responsible for ensuring compliance with the policy framework for internal summative assessment?	Which reference standards are used for internal summative assessment?	What mechanisms are in place to ensure the reliability of marking in internal summative assessment across students (within and between schools)?	What is the weight of internal summative assessment in determining students' year-end marks?
		1	2	3	4	5
Spain	General	Yes, at the central level for all schools[14]	State inspection bodies	National curriculum goals	Availability of national guidance materials for marking student performance on the examination (performance criteria)	100%
	Pre-voc and voc	No	a	a	a	a
Sweden	All programmes	No	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	Yes, at the central level for all schools (Department of Education, Northern Ireland)	Council for Curriculum, Examinations and Assessment (CCEA)	Statutory rule on Levels of Progression which is to be laid before the Northern Ireland Assembly	There is both internal moderation within schools and moderation by CCEA	Not specified in the framework

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

1. Australia: A national curriculum (The Australian Curriculum) is currently being implemented and is expected to replace state-based curriculum frameworks.

2. Belgium (Fl.): At present, the Ministry of Education and Training does not write comprehensive curricula. However, the Ministry specifies attainment targets and developmental objectives (*eindtermen en ontwikkelingsdoelen*), which function as the basis for each curriculum. The attainment targets and developmental objectives are also called the "core curriculum". They are set by the Flemish central education authority. It is mandatory for schools to monitor their educational quality, which implies that each school must at least assess the extent to which students acquire the skills, knowledge and attitudes described in the attainment targets and developmental objectives. In this respect each school is required to develop an assessment policy which includes an output-based monitoring of the achievements of its students. The Inspectorate may ask schools to present their policy. To some extent, internal summative assessment at schools is regulated by the attainment targets. Schools are also required to assess whether students have attained the objectives of the school curriculum to a satisfactory level. However, there are no regulations regarding the methods schools can use to do so.

3. Belgium (Fr.): The policy framework is provided through the work of the Inspectorate and the possibility for students to appeal summative assessment decisions made by the class council through an appeals council.

4. Chile: The policy framework sets very basic requirements, such as the number of assessments per year. It also states that every school has to define an internal protocol for summative assessment.

5. Czech Republic: The central education authority does not determine the framework for the internal summative evaluation. However, the Education Act requires every school to set its own assessment system that is approved by the school board. Internal school assessment regulations and their administration are controlled by the Czech School Inspectorate (Česká školní inspekce [CSI])

6. Denmark: A few private schools may be exempt from these requirements.

7. Finland: To some extent, the national core curriculum functions as a policy framework that regulates internal assessments.

8. Finland: In basic education, the marking scale ranges from 4 (weak) to 10 (excellent) and the curriculum contains descriptions of "good" performance (mark 8) at the so-called transition points for every subject. At the end of basic education there are final assessment criteria for the mark 8. Teachers hold responsibility for the design of assessments, but design assessments in relation to these descriptions and criteria. The tests are, therefore, based on the same criteria.

9. Iceland: At completion of compulsory school (Year 10), students undergo a summative assessment. It is, however, at the discretion of schools themselves how this is implemented and may take the form of a formative assessment. Requirements stipulated in legislation, regulations and curriculum guidelines are flexible.

10. Ireland: Guidance material is available from syllabus assessment guidelines, the State Examinations Commission and other sources. Syllabus and curricular guidelines, developed by the National Council for Curriculum and Assessment, provide assessment supports, as does material provided by the Professional Development Service for Teachers. The State Examinations Commission provides guidance for teachers in summative assessment through the publication of examination papers, marking schemes and chief examiners' reports in different subjects.

11. Italy: Presidential Decree n. 122/2009 provides norms on student assessment. The national curriculum goals are not explicitly mentioned in this regulation, but they are normally considered by school committees.

12. Norway: There are regulations in the Education Act that state that students at this level shall receive summative overall achievement marks in all subjects in addition to ongoing formative assessment and formal formative feedback minimum twice a year. The overall achievement marks based on teachers judgements are entered on the leaving certificate in Year 10.

13. Poland: While a policy framework exists, it is very general and leaves a great degree of autonomy to schools to define their own assessment rules. Schools are, however, required by law to define assessment regulations as part of the school statute.

14. Spain: Subject teachers are responsible for the assessment and marking of stud+A38ents at this level. There are, however, some constraints and guidelines. Minimum subject contents, learning objectives (both summative and formative) and assessment criteria are established at a central level. At the school level subject departments write a general programme for the year, revised by the inspection services to make sure it complies with the legal framework, establishing the assessment and marking methods and criteria. Subject departments are also responsible for making a written test (optional for those students who have previously attained the objectives of the school year), common to all school groups of the same level for year-end assessment and marking. Finally, the class teachers board has a say in the end of year assessment, especially at the end of each cycle and in relation to the attainment of the general objectives of the stage.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A1.2 Student formative assessment frameworks at ISCED levels 1, 2 and 3 (2012)

This table describes central/state policy frameworks for student formative assessment, i.e. assessment that aims to identify aspects of learning as it is developing in order to deepen and shape subsequent learning rather than make a judgement about the past performance of the student. It is essentially a pedagogical approach consisting of frequent, interactive checks of student understanding to identify learning needs and adapt teaching strategies.

Country	Is there a policy framework for promoting student formative assessment in the classroom?	What requirements are part of the policy framework for promoting student formative assessment in the classroom?
	1	2
Australia	ISCED levels 1, 2 and 3 (general): Yes, at the state level for all schools ISCED level 3 (pre-voc and voc): No	ISCED level 1: For schools to implement strategies for student formative assessment ISCED levels 2 and 3: For schools to implement strategies for student formative assessment; for student formative assessment to be part of initial teacher education programmes
Austria	No	a
Belgium (Fl.)	No[1]	a
Belgium (Fr.)	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment
Canada	Varies across provinces/territories	Varies across provinces/territories
Chile	No	a
Czech Republic	No	a
Denmark	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment[2]
Estonia	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment; for schools to report on their strategies to promote student formative assessment
Finland	No[3]	a
France	No	a
Hungary	No	a
Iceland	ISCED levels 1 and 2: yes, at the central level for all schools[4] ISCED level 3: no	For schools to implement strategies for student formative assessment
Ireland	ISCED level 1: yes, at the central level for public schools and government-dependent private schools only ISCED levels 2 and 3: no[5]	For schools to implement strategies for student formative assessment
Israel	ISCED levels 1 and 2: yes, at the central level for all schools ISCED level 3: no	For schools to implement strategies for student formative assessment
Italy	Yes, at the central level for all schools[6]	For schools to implement strategies and criteria for student formative assessment
Korea	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment; for student formative assessment to be part of initial teacher education programmes; for teachers to undertake professional development in student formative assessment

Table 4.A1.2 Student formative assessment frameworks at ISCED levels 1, 2 and 3 (2012) (*continued*)

Country	Is there a policy framework for promoting student formative assessment in the classroom?	What requirements are part of the policy framework for promoting student formative assessment in the classroom?
	1	2
Luxembourg	Yes, at the central level for public schools only	For schools to implement strategies for student formative assessment
Mexico	ISCED 1 and 2: yes, at the central level for all schools ISCED 3: yes, at the central level for centrally managed public schools, at the state level for locally managed public schools[7]	For schools to implement student formative assessment; for student formative assessment to be part of initial teacher education programmes[8]
Netherlands	No[9]	a
New Zealand	Yes, at the central level for state schools and state-integrated schools only	For schools to implement strategies for student formative assessment
Norway	Yes, at the central level for all schools[10]	For schools to implement student formative assessment
Poland	ISCED levels 1, 2 (general), ISCED level 3: yes, at the central level for all schools ISCED level 2 (pre-voc and voc): No	For schools to implement student formative assessment[11]
Portugal	ISCED 1, 2 (general) and 3 (general): yes, at the central level for all schools	For schools to implement strategies for student formative assessment (approved by the pedagogical council at each school)
Slovak Republic	No	a
Slovenia	Yes, at the central level for all schools	For schools to implement student formative assessment
Spain	ISCED levels 1, 2 (general) and 3: yes, at the central level for all schools ISCED level 2 (pre-voc and voc): no	For schools to implement student formative assessment; for schools to report on their strategies to promote student formative assessment; for student formative assessment to be part of initial teacher education programmes; for teachers to undertake professional development in student formative assessment
Sweden	No[12]	a
United Kingdom (Northern Ireland)	Yes, at the central level for all schools	For schools to implement strategies for student formative assessment

Notes: a: information not applicable because the category does not apply; pre-voc and voc: pre-vocational and vocational programmes.

1. Belgium (Fl.): At ISCED level 1, schools are required to monitor the progress of every student and report their observations to parents. Schools are, however, not restricted by any regulations on how to implement progress monitoring or reporting. All relevant agreements on monitoring and assessment have to be included in the school development plan and in the school regulations. At the start of the school year, every parent signs the school regulations for agreement. At ISCED levels 2 and 3, the "counselling class committee" has a formative purpose. The counselling class committee is obliged to monitor the students' performance and assess their progress. The committee is staffed by teachers who teach the students involved. The school principal, deputy principal, technical advisor, members of the Centre for Pupil Guidance (*CLB*) and others may also be asked to participate in meetings of the guiding class committee, e.g. to give advice. According to central government regulations the guiding class committee is not qualified to decide whether a student passes or fails a school year or to take decisive disciplinary measures like refusing students to attend school during the next school year. This decision is made by the deliberative class committee. The counselling class committee must perform the following actions on a regular basis: 1) Analyse and discuss the students' school results as well as their attitudes; 2) Provide students with proper counselling if necessary, e.g. as soon as learning difficulties have been diagnosed; 3) Reduce failure rates by advising students to alter their course of study or – for future reference – to determine the course of study that fits them most.

2. Denmark: Individual Student Plans are compulsory in all subjects at least once a year for Years 1 to 7. For Years 8 and 9 individual student plans are combined with student learning plans.

3. Finland: The Finnish national core curricula for all ISCED levels mention that teachers should observe student's progress, but not how they should do so. It is at the discretion of local education authorities or individual teachers to decide measures for formative assessment.

4. Iceland: There has been increased emphasis on formative assessment *vis-à-vis* summative assessment. A new policy framework for student assessment is being developed with the publication of the revised curriculum guidelines in 2012.

5. Ireland: For ISCED levels 2 and 3 (general programme only), the National Council for Curriculum and Assessment (NCCA), a central agency, has issued guidelines to schools that emphasise the value and uses of formative assessment (Assessment for Learning). These guidelines are not in the form of regulations or statutes. The new Framework for ISCED 2 to be introduced in 2014 will provide for the promotion of student formative assessment in the classroom.

6. Italy: Central legislation requires schools to set strategies and criteria for formative assessment, but the procedures for doing so are within the scope of school autonomy (Law 122/2009, Artt. 1, 2, 3, 4).

7. Mexico: At schools managed by autonomous agencies (public and private), policies for student formative assessment in the classroom at ISCED level 3 are determined at the school level by autonomous and private institutions (e.g. universities).

8. Mexico: This information reflects the Comprehensive Reform of Basic Education (*Reforma Integral de la Educación Básica* [RIEB]) and the Comprehensive Reform of Upper Secondary Education (*Reforma Integral de la Educación Media Superior* [RIEMS]).

9. Netherlands: There are no formal frameworks for the formative assessment of students. However, there are formative assessment systems (e.g. by the Central Institute for Test Development (*Centraal Instituut voor Toetsontwikkeling* [CITO]) in use in primary and, increasingly, secondary education. Also, draft laws are being prepared that will require schools to use formative assessment systems for results-based work in schools.

10. Norway: Principles of formative assessment are part of the regulations in the Education Act.

11. Poland: The central regulations are very flexible and leave the choice of assessment systems to schools. They only prescribe the format of marks used at the end of the school year. They require the formative use of school assessment by defining the formative use of assessment as one of the main functions of student assessment.

12: Sweden: It is at the discretion of local education authorities or individual teachers to decide measures for formative assessment.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012)

This table describes central/state policy frameworks for the reporting of students' summative results to students and parents at ISCED level 2. Summative results in this table comprise both students' results in standardised national examinations and internal summative assessments.

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Australia	General	Yes, at the central level for all schools; Yes, at the state level for all schools[1]	Varies nationally (on average twice a year)	Alphabetical marks; descriptive ratings; qualitative assessment	Varies nationally (on average twice a year)	Mark in each of the concerned subjects
	Pre-voc and voc	Yes, at the state level for all schools (Australian Qualifications Framework)[1, 2]	Twice a year	Descriptive ratings; alphabetical mark	Twice a year	Statement of attainment of competencies
Austria	All programmes	Yes, at the central level for public schools and government-dependent private schools only	Twice a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Belgium (Fl.)	All programmes	No[3]	a	a	a	a
Belgium (Fr.)	All programmes	No	a	a	a	a
Canada	All programmes	Yes, at the provincial/territorial level for public schools and government-dependent private schools only	In most cases once at the end of the school year	Varies across provinces/territories	Varies across provinces/territories	No formal certificate of learning is awarded
Chile	All programmes	Yes, at the central level for all schools	Typically 2-3 times a year	No requirement exists	Typically 2-3 times a year	Primary certificate and marks in each of the concerned subjects
Czech Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Denmark	All programmes	Yes, at central level for all schools	At least twice a year	Numerical marks (-2 to 12)	Once a year	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in student project in Year 9

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Estonia	General	Yes, at central level for all schools	At least twice a year	Schools can decide whether to provide numerical marks or alphabetical marks	Once a year	Mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Finland	All programmes	Yes, at the central level for all schools[4]	Twice a year	Requirement to report numerical marks (4-10) from Year 8 onwards	No requirement exists	Overall mark; mark in each of the concerned subjects
France	All programmes	Yes, at the central level for all schools	At least 3 times a year	Numerical marks (0-20); descriptive ratings of competencies; qualitative assessment	At least three times a year; at the request of parents	Mark in each of the concerned subjects; summary of the student's overall strengths and weaknesses; progression to the next year level; grade point average
Hungary	All programmes	Yes, at the central level for public schools only	Twice a year	Numerical marks (1-5)	No requirement exists[5]	Mark in each of the concerned subjects
Iceland	All programmes	Yes, at the central level for all schools for Year 10	Twice a year	Qualitative assessment; numerical marks (0–10)	No requirements exists	Mark in each of the concerned subjects; summary of the student's strengths and weaknesses in each subject
Ireland	General	No, but advice and guidance are provided by a central agency (National Council for Curriculum and Assessment [NCCA])[6]	Typically twice a year	Typically numerical/ alphabetical marks and descriptive ratings	Typically once a year; Additional meetings facilitated at the request of parents/teachers	Typically alphabetical mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Israel	All programmes	Yes, at the central level for all schools[7]	Twice a year	Descriptive ratings; qualitative assessment; numerical marks	Twice a year	Mark in each of the concerned subjects

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Italy	All programmes	Yes, at the central level for all schools[8]	2-3 times a year	Numerical marks (1-10)	No requirement exists[9]	Mark in each of the concerned subjects; overall mark (provided in ISCED 2 certification at the end of Year 8 based on average of all tests including the national examination)
Korea	All programmes	Yes, at the central level for all schools	4 times a year	Numerical marks (0-100) and rankings	No requirement exists	Pass/fail information
Luxembourg	All programmes	Yes, at the central level for all schools	3 times a year	Numerical marks (0-60 points), descriptive ratings[10]	At the request of students/parents only[11]	Mark in each of the concerned subjects
Mexico[12]	All programmes	Yes, at the central level for all schools	More than 4 times a year	Numerical marks (5-10)	More than 4 times a year	Pass/fail information; overall mark (provided in the *Certificado de Terminación de Estudios*)
Netherlands	All programmes	Yes, at the central level for all schools	3 times a year	Numerical marks (0-10)	3 times a year	Pass/fail information; mark in each of the concerned subjects
New Zealand	All programmes	No	a	a	a	a
Norway	All programmes	Yes, at the central level for all schools	Once a year	Numerical mark (1-6)	Twice a year	Mark in each of the concerned subjects
Poland	General	Yes, at the central level for all schools	Twice a year	Descriptive ratings;[13] mark for behaviour/conduct	No requirement exists	Descriptive ratings;[13] mark for behaviour/conduct; comments on additional coursework and special achievements
	Pre-voc and voc	No	a	a	a	a
Portugal	General	Yes, at central level, for all schools	3 times a year	Numerical marks (1-5)	3 times a year and/or at request of parent	A formal certificate of learning is awarded when requested; mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a

Table 4.A1.3 Reporting of summative results at ISCED level 2 (2012) (*continued*)

Country	Programme type	Is there a policy framework that regulates the reporting of summative results?	How often are student summative results formally reported in writing?	What type of information is provided in writing on student results?	How often do teachers have to hold formal summative feedback meetings with students/parents?	What type of written information is provided to students at the end of ISCED level 2 on formal certificates of learning?
		1	2	3	4	5
Slovenia	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5)	Twice a year	Mark in each of the concerned subjects
Slovak Republic	All programmes	Yes, at the central level for all schools	Twice a year	Numerical marks (1-5) or descriptive ratings	3-4 times a year	No formal certificate of learning is awarded
Spain	General only	Yes, at the central level for all schools[14]	3 times a year	Numerical marks (0-10)	At the request of parents	Mark in each of the concerned subjects
	Pre-voc and voc	No	a	a	a	a
Sweden	All programmes	Yes, at the central level for all schools	Twice a year	Descriptive ratings; qualitative assessments	Twice a year	Mark in each of the concerned subjects
United Kingdom (Northern Ireland)	All programmes	Yes, at the central level for all schools	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[15]	Once a year	The level of progression achieved in Communication and Using Maths (from 2012/13)[15]

Notes: a: information not applicable because the category does not apply; pre-voc and voc: pre-vocational and vocational programmes.

1. Australia is a federation of eight states/territories. Standardised examinations and internal summative assessment practices are set at the state/territory level and there is a national requirement to report student results in a nationally consistent way (A-E reporting).

2. Australia: Vocational education and training (VET) qualifications approved under the Australian Qualifications Framework (AQF) at ISCED levels 2 and 3 are typically offered by the VET sector. Students may take VET qualifications as part of a general programme (e.g. through the programme VET in Schools). VET (AQF) qualifications are delivered by a Registered Training Organisation (RTO) or a school in partnership with a RTO. All schools (public and private) are able to undertake partnerships with RTOs to provide students with an opportunity to undertake VET courses.

3. Belgium (Fl.): In secondary education, the "deliberative class committee" decides whether or not students pass or fail. The deliberative class committee's decisions rely on concrete data from the file of the student. The "counselling class committee" has a formative purpose. The counselling class committee is obliged to monitor the students' performance and assess their progress. The information in primary and secondary education can be provided as a qualitative assessment or by numerical marks (overall mark and marks/subject), depending on the school policy. At the end of Years 2 and 4 of lower secondary education, students receive a certificate.

4. Finland: The information provided refers to the reporting of summative results through report cards/certificates. Assessment is based on the national core curriculum. Standardised national examinations are only held at the end of ISCED level 3 (general programme).

5. Hungary: At the discretion of each school.

6. Ireland: Advice and guidance are provided by the National Council for Curriculum and Assessment (NCCA), but not in form of a statute.

7. Israel: The information provided refers mainly to a report card at the end of the school year.

8. Italy: From age 16 onwards, ISCED level 2 qualifications can also be obtained from provincial centres for adult education.

9. Italy: Decisions on formal summative feedback meetings with students/parents rest within schools, per law n. 122/2009. Typically meetings take place 2-3 times a year.

10. Luxembourg: For details, see the following website: *www.men.public.lu/priorites/ens_fondamental/090723_bibliotheque/111201_intermediate_reports_cycle2.pdf*.

11. Luxembourg: Each secondary school is required to organise an information meeting once a year (1st or 2nd trimester). However, there is no obligation to provide summative feedback as part of these meetings.

12. Mexico: From 2012-13 onwards, the new general norms on evaluation, accreditation, promotion and certification in basic education will apply. The Basic Education Certificate (*Certificado de Educación Básica*) will be issues on completing Year 12.

13. Poland: 6: Excellent, 5: Very good, 4: Good, 3: Satisfactory, 2: Acceptable, 1: Unsatisfactory.

14. Spain: More complete reports are given only to students with special educational needs or when a learning problem arises. This typically involves the identification of learning difficulties and related advice (e.g. year repetition, curricular adaptation, etc.).

15. United Kingdom (Northern Ireland): From 2013/14, the levels of progression achieved in Using ICT will also be provided.

Source: Derived from information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central examinations

Austria: The introduction of national examinations at ISCED level 3 starts in school year 2014/15 for general programmes, and in school year 2015/16 for pre-vocational and vocational programmes. A central agency, the Federal Institute for Research, Innovation and Development of the Austrian School System (*Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens* [BIFIE]) is responsible for the development of the central examination. The following subjects will be examined in a standardised form at ISCED level 3: the language of instruction (German, Croatian, Hungarian, Slovenian), foreign languages (English, French, Spanish, Italian, Latin, Greek) and mathematics. Schools/teachers will be responsible for marking the new national examination from 2014/15 onwards following guidelines developed by BIFIE.

Belgium (Fr.): Standardised central examinations at ISCED levels 2 and 3 are not mandatory. Schools decide about their administration. At ISCED level 2, 55.6% of students participated in these examinations in 2011/12. At ISCED level 3, standardised national examinations in general programmes examine one competency area in history, and standardised examinations in pre-vocational and vocational programmes examine one competency area (reading) in the language of instruction (French). Schools decide about the weight given to standardised examination results in relation to other results of the student. In 2011/12, 42.9% of students participated in these examinations.

Iceland: With the introduction of the Compulsory School Act 2008, standardised national examinations were changed to standardised national assessments.

New Zealand: Students can participate in national qualifications examinations at any stage and it is not uncommon for some students to participate at ISCED level 2. However, the majority of the students undertake the national qualifications at ISCED level 3.

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012)

This table describes central/state policy frameworks for standardised central examinations, i.e. standardised student assessments that have a formal consequence for students (e.g. impact upon a student's eligibility to progress to a higher level of education, part of a process to certify learning) for ISCED level 2.

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Australia	All programmes	No	a[1]	a	a	a	a	a	a	a	a	a	a
Austria	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fl.)	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Belgium (Fr.)	General	Yes	*CE1D (Épreuve certificative externe commune au terme de la troisième étape du continuum pédagogique)*	Yes, for public schools only (from 2012/13)	All students: M, L	Year 8	Central education authority	School[2]	Availability of national guidance materials for marking student performance on the examination	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National standards
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Canada	All programmes	Yes, at the provincial level	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province	Varies by province
Chile	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Czech Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Denmark	All programmes	Yes	*Folkeskolens Afsluttende Prøver*	Yes, for public schools only[3]	All students: M, L, S, FL Students choose: A, V Sample: FL, S, SS, R	Year 9	Central education authority	Centrally appointed external examiners mark student performance in co-operation with internal examiners	Availability of national guidance materials for marking student performance on the examination (performance criteria, exemplars, rubrics, keys); moderation of marking (external examiners attend all oral examinations at school level)	Multiple choice; Open-ended writing tasks; oral presentation; oral questions and answers; project presentation; group discussion (optional)	Closed-format short-answer questions, open-ended writing tasks	No	National learning progressions

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Estonia	General	Yes	Basic school end exams	Yes, for public and government-dependent private schools only	All students: M; L Students choose: S; SS; FL	Year 9	State education authorities or governments; state agency responsible for assessment or certification	The student's own teacher	Moderation of marking; external checking of a sample of student products by a competent body; availability of national guidance materials for marking student performance on the examination (performance criteria, rubrics, exemplars)	Multiple choice; closed format short-answer questions; open-ended writing tasks	Performing tasks	No	National curriculum goals
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Finland	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
France	All programmes	Yes	*Diplôme national du brevet (DNB)*[4]	Yes, for all schools	All students: M, L, S, SS, FL, T, A, R	Year 9	Central education authority	Continuous assessment: the student's own teacher. For the standardised part of the examination: a teacher from another school	None	Open-ended writing tasks (essays); continuous assessment	Open-ended writing tasks (essays); continuous assessment	No	National curriculum goals
Hungary	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Iceland	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Ireland	All programmes	Yes	Junior Certificate	Yes, for public and government-dependent private schools only	All students: M, L, S, SS Students choose: A, R, V, O	Year 11	Central agency responsible for assessment (State Examination Commission)	Central agency responsible for assessment (State Examination Commission)	a	Open-ended writing tasks; closed-format short-answer questions	Written tasks (Closed-format tasks; open-ended tasks; general problem solving; problem-solving in unfamiliar contexts)	No	National curriculum goals
Israel	All programmes	No	a	a	a	a	a	a	a	a	a	a	a

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Italy	General	Yes	*Esame di Stato conclusivo del primo ciclo di istruzione (Prova nazionale)*	Yes, for all schools	All students: M, L, S, FL, A	Year 8	Central education authority (INVALSI) for the standardised central examination, which is a portion of the national 1st cycle examination	School (School Examination Committee) on the basis of INVALSI correction grids; Central agency responsible for assessment (National Institute for Assessment [INVALSI])[5]	a[5]	Multiple choice; closed-format short-answer questions; open-ended writing tasks. School-based part of the examination: written essay; oral questions and answers; oral presentation	Multiple choice; closed-format short-answer questions; open-ended writing tasks. School-based part of the examination: written problems/exercises; oral questions and answers; oral presentation	No	National curriculum goals
Korea	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Luxembourg	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Mexico	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Netherlands	General	Yes	Examination in vmbo-t programmes	Yes, for all schools	All students: L, FL, SS Students choose: M, S, FL, O	Year 12	Central agency (Central Institute for Test Development [CITO])[6]	The student's own teacher	Availability of national guidance materials; moderation of marking (systematic external moderation by school organising bodies [competent authorities])[7]	Multiple choice; open-ended writing tasks	Open-ended calculations	No	National curriculum goals
	Pre-voc and voc	Yes	Examination in vmbo-b, vmbo-k, vmbo-gt programmes	Yes, for all schools	All students: L, FL, V, SS Students choose: M, S, FL, O	Year 12	Central agency (Central Institute for Test Development [CITO])[6]	The student's own teacher	Availability of national guidance materials; moderation of marking (systematic external moderation by school organising bodies [competent authorities])[7]	Multiple choice; open-ended writing tasks	Open-ended calculations	Yes	National curriculum goals
New Zealand	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Norway	All programmes	Yes	*Sentralt gitt eksamen* (National exam)	Yes, for all schools	Sample: M, L, FL[8]	Year 10	Central education authority	Central education authority	a	Open-ended writing tasks	Closed-format short-answer questions; open-ended writing tasks	Yes, computer-based uniform technology (in mathematics only for part two of the exam)	National curriculum goals[9]

Table 4.A1.4 Standardised central examinations at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central examinations exist at the lower secondary level?	Name of the standardised central examination at ISCED level 2	Are central examinations compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central examination?	Who is responsible for marking the central examination?	If marking of central examinations (or of parts of the examinations) is undertaken at the school level, what mechanisms are in place to ensure the reliability of marking across students (within and between schools)?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the examinations?	Which reference standards are used for the marking of standardised central examinations in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11	12
Poland	General	Yes[10]	m	Yes, for all schools	All students: M, L, S, SS, FL	Year 9	Central and regional agencies responsible for external assessment (Central Examination Commission, Regional Examination Commissions)	Central and regional agencies responsible for external assessment (Central Examination Commission, Regional Examination Commissions)	a	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	National examination standards based on core curriculum[11]
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Portugal	General	Yes	Final national examinations	Yes, for all schools	All students: M, L	Year 9	Central education authority	Teacher from another school	Availability of national guidance materials (performance criteria); moderation of marking (teachers discussing student performance in groups)	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks; problem-solving tasks	No	National curriculum (performance criteria)
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a	a
Slovak Republic	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Slovenia	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Spain	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
Sweden	All programmes	No	a	a	a	a	a	a	a	a	a	a	a
United Kingdom (Northern Ireland)	All programmes	No	a	a	a	a	a	a	a	a	a	a	a

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Australia: There are no standardised central examinations at ISCED level 2 in Australia. However, most schools have some sort of mandatory assessments in Year 10 level. These assessments largely do not have an effect upon students' progression to a higher level of education or completion of an officially recognised degree.

2. Belgium (Fr.): Schools organise themselves for the marking of standardised central examinations. This may take different forms, e.g. correction by the teacher, another teacher, in groups, etc.

3. Denmark: 95% of government-dependent private schools also administer standardised central examinations, although they are not required to do so.

4. France: The DNB is based 60% on continuous assessment by teachers and 40% on a standardised examination marked by teachers from other schools.

5. Italy: The National Institute for Assessment (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI]) marks a standardised test worth 1/7 of the mark (which is an average of mark in national Invalsi examination, marks in non national assessments, a mark in oral colloquium and a mark in admission to the examination). There is a school-based, non standardised part of the examination which is scored locally. No regulations exist to date to moderate student marks for the non standardised part of the examination.

6. Netherlands: *Centraal Instituut voor Toetsontwikkeling.*

7. Netherlands: In a first step, marking guidelines set by a central agency, the Central Committee for Examinations, are used by the student's own teachers. In a second step, a second examiner, a teacher from another school, examines student examinations. In case of disagreement, the organising body (competent authority) of the second external examiner notifies the organising body (competent authority) of the first examiner (the student's own teacher). In case the disagreement cannot be resolved, the Inspectorate intervenes. The Inspectorate can appoint a third independent examiner that takes the ultimate decision about a student's mark.

8. Norway: All students in Year 10 sit a centrally given written examination in one subject (Norwegian, mathematics or English) and a locally given oral examination in one subject. Students are sampled randomly for the different subjects.

9. Norway: As specified in the curricula, the education law and specific subject assessment guidelines for examinations.

10. Poland: Standardised examinations at ISCED level 2 cover mathematics, science, the language of instruction, modern foreign languages and social studies. Since 2012, scores are reported separately for each subject. Between 2002 and 2011 scores were combined for mathematics and science and for the language of instruction and social studies. Scores for modern languages were already reported separately.

11. Poland: From 2012 onwards, national examination standards have been replaced by the new core curriculum, which is formulated in terms of learning outcomes.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

General notes on standardised central assessments

Belgium (Fr.): Standardised central assessments (*Évaluation externe des acquis des élèves de l'enseignement obligatoire*) are conducted for the purpose of identifying individual learning needs. They are compulsory for all students in the given years. Subjects are rotated on the basis of 3-year cycles (mathematics was tested in 2011/12).

Norway: Standardised central assessments examine competencies in reading literacy and numeracy. There are no tests in the language of instruction, modern foreign languages or mathematics.

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012)

This table describes central/state policy frameworks for standardised central assessments, i.e. standardised student assessments which are mostly used to monitor the quality of education at the system and/or school level, at ISCED level 2.

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Australia	General	Yes	National Assessment Program – Literacy and Numeracy (NAPLAN); National Assessment Program – Civics and Citizenship (NAP-CC); National Assessment Program – Information and Communication Technology Literacy (NAP-ICTL)	Yes, for all schools	All students: M (Years 7, 9); L (Years 7, 9) Sample: SS (Year 10 only); T (Year 10 only)	Years 7, 9, 10	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Central agency responsible for assessment (Australian Curriculum, Assessment and Reporting Authority [ACARA])	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions	No, except computer based uniform technology is used for students with special needs and for NAP-ICTL	National learning progressions
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Austria	All programmes	Yes	*Bildungsstandards* (Educational Standards)	Yes, for public and government-dependent private schools only	All students: M, L, FL	Year 8	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[1]	Central agency responsible for assessment (Federal Institute for Education Research, Innovation and Development of the Austrian School System [BIFIE])[1]	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation; oral questions and answers[2]	Multiple choice; closed-format short-answer questions; open-ended writing tasks[2]	No	National standards

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Belgium (Fl.)	All programmes	Yes	National Assessment Programme	No	Sample: varies[3]	Varies[3]	Research institute commissioned by central education authority	Research institute commissioned by central education authority	Multiple choice; closed-format short-answer questions; performing a task/experiment	Multiple choice; closed-format short-answer questions; performing a task/experiment	Yes, computer-based uniform technology for some tests	National standards
Belgium (Fr.)	General	Yes	*Évaluation externe non certificative*	Yes, for all schools	All students: varies[4]	Varies[4]	Central education authority	School	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task[4]	Multiple choice; closed format short-answer questions; open-ended writing tasks; performing a task/ experiment	No	National standards[4]
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Canada	All programmes	Yes	Pan-Canadian Assessment Program (PCAP)	Yes, for public and government-dependent private schools only	Sample: M, L, S[5]	Year 8	Central education council (Council of Ministers of Education, Canada)	Central education council (Council of Ministers of Education, Canada)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	Pan-Canadian Assessment Framework
Chile	All programmes	Yes	System for Measuring the Quality of Education (SIMCE)[6]	Yes, for all schools	All students: M, L, S, SS (Year 8, every 2 years) Sample: O (physical education) (Year 8)	Year 8	Central education authority (Ministry of Education)[7]	Central education authority (Ministry of Education)[7]	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No	National curriculum goals; national standards
Czech Republic	All programmes	No, but central assessments are currently being piloted[8]	Central assessment	Yes, for all schools	All students: M, L, FL	Year 9	Czech School Inspectorate (CSI)[9]	Czech School Inspectorate (CSI)[9]	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National curriculum goals; national standards

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Denmark	All programmes	Yes	*Nationale test*	No	All students: M (Year 6), L (Years 6, 8), FL (Year 7 only), S, SS (Year 8)	Varies by subject	Central agency responsible for assessment (National Agency for Quality and Supervision)	Central agency responsible for assessment (National Agency for Quality and Supervision)	Multiple choice; closed-format writing tasks; matching items (e.g. pictures/drawings with words)	Multiple choice; closed-format writing tasks; matching items	Yes, computer-based adaptive technology[10]	National learning progressions
Estonia	All programmes	No	a	a	a	a	a	a	a	a	a	a
Finland	All programmes	Yes[11]	Sample based assessments of learning outcomes	No	Sample: Subjects vary, but typically include M, L; other subjects assessed include S, FL, A, R, and cross-curricular learning-to-learn skills[9]	Varies (most regularly Year 9, also Year 7)	Central education authority (Finnish National Board of Education)	The student´s own teacher (according to central marking guidelines)	Multiple choice; closed-format short-answer questions; open-ended writing tasks; oral presentation	Multiple choice; closed-format short-answer questions	No	Marking guidelines
France	All programmes	Yes	*Évaluations-bilan CEDRE (cycle des évaluations disciplinaires réalisées sur échantillon)* *LOLF-base (indicateurs liés à la loi organique relative à la loi des finances)*	Yes, for all schools	*Évaluations-bilan CEDRE*: Sample: A different discipline each year (6-year cycle) *LOLF-base*: Sample: M, L	Year 9	Central education authority (General Directorate for School Education [DGESCO]);[12] central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[13]	Central agency responsible for assessment (Directorate for Evaluation, Forecasting and Performance [DEPP])[13]	Multiple choice; short open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National curriculum goals; national standards
Hungary	All programmes	Yes	National Assessment of Basic Competencies	Yes, for all schools	All students: M, L	Years 6, 8	Central education authority	Central education authority	Multiple choice; short open-ended writing tasks	Multiple choice; short open-ended writing tasks	No	National Assessment Framework
Iceland	All programmes	Yes	m	Yes, for all schools	All students: M, L, S, FL	Year 10	Central agency responsible for assessment (Educational Testing Institute)	Central agency responsible for assessment (Educational Testing Institute)	Multiple choice; open-ended writing tasks; oral questions and answers	Multiple choice; open-ended writing tasks; oral questions and answers	No	National curriculum goals
Ireland	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Israel	All programmes	Yes	*Meitzav*	Yes, for all schools according to cycles	All students: M, L, S, FL	Year 8	Central education authority or government	Central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals
Italy	General	Yes	*Servizio Nazionale di Valutazione (SNV)*	Yes, for all schools	All students: M, L	Years 6, 8	Central education authority with support of central agency responsible for assessment (National Institute for Assessment [INVALSI])[14]	Central education authority with support of central agency responsible for assessment (National Institute for Assessment [INVALSI])[14]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice; closed-format short-answer questions; open-ended writing tasks	No	National curriculum goals[14]
Korea	All programmes	Yes	Subject Learning Diagnostic Test; National Assessment of Educational Achievement (NAEA)	Yes, for all schools	Subject Learning Diagnostic Test: All students: M, L, S, SS NAEA: All students: M, L, S, SS, FL	Subject Learning Diagnostic Test: Years 7, 8, 9 NAEA: Years 6, 9	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Central agency responsible for assessment (Korea Institute for Curriculum and Evaluation [KICE])	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	No	National curriculum goals
Luxembourg	All programmes	Yes	*Épreuves standardisées (5ème ES / 9 EST)*	Yes, for public schools only	All students: M, L (German, French)	Year 8, 10	Central education authority; research institute (University of Luxembourg)	research institute (University of Luxembourg)[15]	Multiple choice	Multiple choice	Yes, computer-based uniform technology	National standards
Mexico	All programmes	Yes	*ENLACE Básica*	Yes, for all schools	All students: M, L; further subject on rotating annual basis (e.g. S, SS)	All students: Years 7, 8, 9 Sample: Year 9[16]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation ([INEE])[17]	Central education authority; central agency responsible for assessment (National Institute for Educational Assessment and Evaluation [INEE])[17]	Multiple choice; closed-format short-answer questions; open-ended writing tasks	Multiple choice	No	National curriculum goals
Netherlands	All programmes	No	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
New Zealand	All programmes	Yes	National Education Monitoring Project	No	Sample: 4 blocks of curriculum areas assessed alternately on a 4-year cycle: (1) S, A, O (information skills [graphs, tables, maps, charts, diagrams]); (2) L, T, A; (3) M, SS, O (information skills [library, research]); (4) L, O (health and physical education)	Year 8	Central education authority or government; Private company contracted to Central education authority or government	Central education authority or government; Private company contracted to central education authority or government	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Multiple choice; closed-format short-answer questions; open-ended writing tasks; performing a task/experiment; oral questions and answers	Yes, computer-based uniform technology	National curriculum goals
Norway	All programmes	Yes	National tests	Yes, for all schools	All students: M, L, FL	Years 8, 9	Central education authority (Norwegian Directorate for Education and Training)	The student's own teacher; central education authority (Norwegian Directorate for Education and Training)	Multiple choice; closed-format short-answer questions	Multiple choice; closed-format short-answer questions	Yes, computer-based uniform technology	National learning progressions
Poland	All programmes	No	a	a	a	a	a	a	a	a	a	a
Portugal	All programmes	No	a	a	a	a	a	a	a	a	a	a
Slovak Republic	General	Yes	*Testovanie 9*	Yes, for all schools	All students: M, L	Year 9	Central agency responsible for assessment (NÚCEM)	Central agency responsible for assessment (NÚCEM)	a	Multiple choice; closed-format short-answer questions	m	National curriculum goals
	Pre-voc and voc	No	a	a	a	a	a	a	a	a	a	a
Slovenia	All programmes	Yes	National Assessment	Yes, for all schools	All students: L, M, O (subject determined by the Ministry for Education and Sport)	Year 9	Central agency responsible for assessment (National Examinations Centre)	Central agencies (National Examinations Centre; National Education Institute)	Multiple choice; open-ended writing tasks	Multiple choice; open-ended writing tasks	No, computer-based technology for some students with special needs only[18]	National curriculum goals
Spain	All programmes	No[19]	a	a	a	a	a	a	a	a	a	a

Table 4.A1.5 Standardised central assessments at ISCED level 2 (2012) (*continued*)

Country	Programme type	Do central assessments exist at ISCED level 2?	Name of central assessment at ISCED level 2	Are central assessments compulsory?	What subjects are assessed?	Which year levels are assessed?	Who is responsible for developing the central assessment?	Who is responsible for marking the central assessment?	Which types of tasks do students have to complete in the language of instruction?	Which types of tasks do students have to complete in mathematics?	Is computer-based technology used for the administration of the assessments?	Which reference standards are used for the marking of standardised central assessments in the language of instruction and mathematics?
		1	2	3	4	5	6	7	8	9	10	11
Sweden	All programmes	Yes	National tests	Yes, for all schools	All students: M, L, FL, S	Year 9	Central educational authority or government; central agency (National Agency for Education)	The student's own teacher (another teacher from within the school may collaborate; Inspectorate reviews the marking of a sample of tests)	Multiple choice; open-ended writing tasks; oral presentation	Multiple choice; closed-format short-answer questions; oral presentation; oral questions and answers	No	National knowledge requirements
United Kingdom (Northern Ireland)	All programmes	Yes	Assessments using Levels of Progression (legislation is currently in train)	Yes, for all schools (from 2012/13)	All students: Communication and Using Maths[20]	Year 10 – level 5 expected	Department of Education tasks the Council for Curriculum, Examinations and Assessment (CCEA) to develop the assessments	Department of Education tasks CCEA to develop the assessments	A range of teacher-set tasks including some tasks set by CCEA	A range of teacher set tasks including some tasks set by CCEA	Yes	Levels of Progression as set out in legislation and subsequently in guidance developed by CCEA

Notes: a: information not applicable because the category does not apply; m: information not available; pre-voc and voc: pre-vocational and vocational programmes.

M: mathematics; L: national language or language of instruction; S: science; SS: social studies; FL: modern foreign languages; T: technology; A: arts; R: religion; V: practical and vocational skills; O: other

All students: all students take the test; Students choose: students can choose to take the test; Sample: sample or selection of students take the test.

1. Austria: *Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens*.

2. Austria: From school year 2011/12 onwards.

3. Belgium (Fl.): There is no systematic assessment of particular learning areas or educational levels. At ISCED level 2 participation in standardised central assessments varies significantly among students in different programmes of secondary education.

4. Belgium (Fr.): Each year, the central assessments cover a different subject (mathematics, language of instruction or science [*éveil-initiation scientifique*]). Subjects are rotated on the basis of a 3-year cycle (e.g. mathematics in 2011/12). In 2011/12, the language of instruction was not assessed.

5. Canada: Mathematics, the language of instruction and science are assessed on a 3-year cycle focussing on one of these three subjects.

6. Chile: *Sistema de Medición de Calidad de la Educación.*

7. Chile: The National Agency for the Quality of Education (*Agencia de Calidad*) will take over these responsibilities from the Ministry of Education in October 2012.

8. Czech Republic: Information provided for the Czech Republic refers to the standardised central assessments being piloted in the academic year 2011/12.

9. Czech Republic: *Česká školní inspekce.*

10. Denmark: Standardised central assessments and school leavers' final examination after Years 9/10 are computer-based. Marking and the calculation of test scores is performed automatically.

11. Finland: The Ministry of Education and Culture has formulated a plan for assessing learning outcomes in basic education for the years 2012-15. Subjects are assessed according to the selection and timings of this plan. At the lower secondary level, a central assessment in mathematics was organised in Year 9 in 2011 and in 2012. An assessment in the language of instruction was conducted in Year 9 in 2010.

12. France: *Direction générale de l'enseignement scolaire.*

13. France: *Direction de l'évaluation, de la prospective et de la performance.*

14. Italy: The National Institute for Assessment (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione [INVALSI]) has developed frameworks for central student assessments in the language of instruction and mathematics which are based on the national curriculum goals.

15. Luxembourg: The tests are automatically marked by a computer.

16. Mexico: Assessed on a 4-year cycle.

17. Mexico: *Instituto Nacional para la Evaluación de la Educación.*

18. Slovenia: From 2013 onwards, the marking of standardised central assessments at ISCED level 2 will be computer based.

19. Spain: Currently, there are no standardised central assessments in Spain. The General Diagnosis Assessment was discontinued in ISCED 2 after 2010. The Spanish government is currently preparing a new education law and a different assessment design for the General Diagnosis Assessment to be re-introduced in the future in line with the new education programme. There are also regional assessment programmes undertaken under the responsibility of the Autonomous Communities.

20. United Kingdom (Northern Ireland): From 2013/14, Using ICT will also be assessed.

Source: Some of the information presented in this table is based on data provided through OECD *Education at a Glance 2011*. This information has been validated and additional information supplied by countries participating in the project. The table should be interpreted as providing broad indications only, and not strict comparability across countries.

Annex 4.A2 Features of student assessment frameworks in primary and upper secondary education

The tables below provide information on features of student assessment frameworks in primary and upper secondary education in the countries actively engaged in the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Part of the information was supplied by countries through a questionnaire specifically developed by the OECD Review. The rest of the information is based on data provided by OECD *Education at a Glance 2011*.

This annex is available on line only at *http://dx.doi.org/10.1787/9789264190658-8-en*. The online annex includes the following material:

Table 4.A2.1a Internal summative assessment frameworks at ISCED level 1 (2012)

Table 4.A2.1b Internal summative assessment frameworks at ISCED level 3 (2012)

Table 4.A2.3a Reporting of summative results at ISCED level 1 (2012)

Table 4.A2.3b Reporting of summative results at ISCED level 3 (2012)

Table 4.A2.4a Standardised central examinations at ISCED level 1 (2012)

Table 4.A2.4b Standardised central examinations at ISCED level 3 (2012)

Table 4.A2.5a Standardised central assessments at ISCED level 1 (2012)

Table 4.A2.5b Standardised central assessments at ISCED level 3 (2012)

All the tables summarising features of evaluation and assessment frameworks, included in the annexes to this report, are also available on the OECD Review website at *www.oecd.org/edu/evaluationpolicy*.

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