Chapter 3

Student assessment

Student performance in Mexico is assessed by a wide range of instruments, ranging from national standardised assessments to continuous formative assessment in the classroom. All students are assessed in an on-going manner throughout the school year in each curriculum area or subject. Marks used to report student achievement are on a scale of 5 to 10. Assessment criteria and methods are defined by each teacher. There are also externally-based national final examinations at the end of both primary (Instrument for Testing New Lower Secondary School Students, IDANIS) and lower secondary education (National Upper Secondary Education Entrance Exam, EXANI I). These assessments serve diagnostic and selection (by school at the next level) functions. At the national level, there is also a full-cohort external assessment (National Assessment of Academic Achievement in Schools, ENLACE) which is used for diagnostic and improvement purposes but which has "high stakes" for teachers and schools. In basic education, ENLACE is administered annually to all students in third to ninth grades in Spanish and mathematics and a third subject which varies every year.

A major asset is that assessment is seen as part of the professional role of teachers in Mexico. Other strengths include the introduction of a new comprehensive framework for classroom-based assessment; the progress made in aligning marks with expected learning outcomes; the good attention to reducing grade repetition; the promotion of the involvement of parents in their children's learning; and the capacity for implementing large-scale assessments. However, considerable challenges exist in building effective student assessment approaches. These include the currently traditional approaches to teaching and assessment; the prevalence of teaching to the test across the school system; the excessive reliance on multiple-choice tests; the great number of objectives for ENLACE; marking practices with little pedagogical significance; the lack of consistency of student assessment across schools and classes; the limited capacities at the state and local levels to support classroom-based assessment; and the need to improve instruments for reporting marks.

This chapter focuses on approaches to student assessment within the Mexican evaluation and assessment framework. Student assessment refers to processes in which evidence of learning is collected in a planned and systematic way in order to make a judgment about student learning (EPPI, 2002). This chapter looks at both summative assessment (assessment *of* learning) and formative assessment (assessment *for* learning) of students.

Context and features

Overview

Student assessment in Mexico comprises three main components: *(i)* classroom-based assessment, with both formative and summative purposes; *(ii)* external assessments for diagnostic and selection purposes both at the entrance of lower secondary education (IDANIS) and of upper secondary education (EXANI I); and *(iii)* external assessments for diagnostic and improvement purposes (ENLACE).

Classroom-based assessment for formative and summative purposes is carried out by teachers and regulated by official Agreement 200 (DOF, 1994). This norm has been in force since 1994 and establishes the obligation for public and private schools to implement student assessment for formative and certification purposes, based on the knowledge, abilities and attitudes included in the national curriculum. This agreement regulates the periodicity of summative assessments, the scale to be used for assigning marks and the conditions to pass or fail a student. In recent years, in the context of the ACE and the RIEB, a new approach to classroom-based assessment has been launched and the new initiatives are in their initial stages of implementation (including proposals to adjust Agreement 200).

IDANIS (Instrument for the Diagnostic of New Lower Secondary School Students, *Instrumento para el Diagnóstico de Alumnos de Nuevo Ingreso a Secundaria*) and EXANI I (National Upper Secondary Education Entrance Examination, *Examen Nacional de Ingreso a la Educación Media Superior*) are national external examinations for diagnostic and selection purposes. IDANIS was created in 1989, is operated by the SEP and assesses students entering lower secondary education. EXANI I was created in 1994, is operated by CENEVAL and assesses students entering upper secondary education. Both instruments are administered with the purpose of providing information to authorities and schools for student selection purposes (Vidal, 2009).

ENLACE is a national standardised assessment administered annually by the SEP on a census basis since 2006. Every student in third to sixth grades of primary education and in seventh to ninth grades of lower secondary education is currently assessed every year in Spanish and mathematics. Since 2008 a third variable subject is assessed each year in all the same grades: sciences in 2008, civics in 2009, history in 2010 and geography in 2011. Results are available individually to each student and average scores by school are widely disseminated. ENLACE's initial explicit purpose was to provide information about students' academic achievement on curriculum objectives, in order to promote the improvement of teachers' professional practices and enhance the quality of learning. Afterwards, new objectives and consequences became attached to these assessments (see below) (SEP and INEE, forthcoming).

Regulations for classroom-based assessment

Classroom-based assessment is regulated by official Agreement number 200 (DOF, 1994). It states that student assessment should be formative, systematic, continuous and integral. It also specifies that for summative purposes teachers must use a numeric scale ranging from 5 to 10. In this scale, 6 is the minimum passing mark, 5 means insufficient and 10 is excellent.

Marks must be assigned and parents informed about them every two months, during October, December, February, April and within the last five working days of the school year, as stated in Agreement 499 (DOF, 2009), which introduced minor amendments to Agreement 200. This norm is widely observed by schools and teachers, leading to an organisation of teaching and learning activities in five terms or blocks, each of them with a two-month duration. School directors must communicate marks to students and parents and foster communication between them and teachers. For this purpose each term schools must complete a standardised report card for each student showing the marks obtained in each subject, which must be signed by parents. The current report card's design corresponds to the 1993 National Curriculum. In the context of the RIEB, a new report card called Basic Education Card is being designed and trialled in schools (see below).

At the end of the school year a final mark on every subject must be calculated as the average of term marks obtained during the school year. Each student also receives an Annual General Average, which is the average of the final marks in each subject (SEP, 2011a).

In primary education there should be no repeating between the first and the second grades, as both are considered part of a unique period of learning. Between second and sixth grades, to pass to the next grade students must obtain at least 6.0 as the final mark in Spanish and mathematics, as well as an Annual General Average equal or greater than 6.0. Students who fail to obtain these marks must repeat the grade, unless they attend "regularisation processes", which are summer courses or tutorships organised and delivered at the state level. Students may pass to the following grade if they are certified by an extraordinary examination before the beginning of the next school year (SEP, 2011a; SEP and INEE, forthcoming).

In lower secondary education students who fail to obtain at least the 6.0 mark in more than five subjects must repeat the whole grade. Students who fail five or fewer subjects at the end of the school year should take development activities and pass an examination in each of those failed subjects. Students cannot enrol in the next grade if they have still not passed more than two subjects after the "regularisation period" in September (at the beginning of the school year) (SEP, 2011a; SEP and INEE, forthcoming).

Classroom-based assessment practices

The procedures and regulations described above are focused on formal aspects of summative assessment. Formative assessment is declared as important, but there are no concrete guidelines about how to perform it. As for summative assessment, the Agreement 200 does not include guidelines about the meaning of marks in terms of expected student performance. This issue is beginning to be addressed in the context of the RIEB, the current curricular reform, including with proposals to revise Agreement 200 (see below).

Within this formal framework, schools and teachers are completely free to determine assessment criteria and undertake student assessment. Teachers use quite different criteria

to assess their students, particularly to assign marks for academic achievement. As observed by the OECD Review Team during the Review visit, many teachers use a normative approach to marks, meaning that they first assign the maximum mark (10) to the best students in their classroom, and then give marks to other students in relation to this benchmark. There are also instances, conveyed to the OECD Review Team by different school agents, where teachers adjust their assessment criteria in order to reduce the number of apparently underachieving students.

For most teachers in Mexico marking consists of assigning points to students across a range of elements: homework completion, class attendance, participation in classroom activities, neatness of tasks, discipline, teamwork, presentations and tests (usually in a multiple-choice format). These aspects receive "points" which are finally averaged to obtain a mark. When asked about which element is the most important for assigning the bi-monthly mark (other than test results), 71% of primary school teachers answered "attention and participation in classroom activities" (García *et al.*, 2011).

The weight assigned to each of these elements varies across teachers and across terms for the same teacher, depending on the content taught and the activities carried out. The weighting of these elements is not explicit, but tests usually have more influence on final marks (García *et al.*, 2009; Loureiro, 2009; Picaroni, 2009; Ravela, 2009a; SEP and INEE, forthcoming).

Formative assessment in Mexico is performed in a very narrow manner: it basically involves giving marks for tests, tasks and attitudes and telling students where they have failed. There is little evidence of teachers' awareness of the importance of giving feedback to students during the process of their work in order to help them reflect about their own learning or about their products (Picaroni, 2009; Ravela, 2009a).

Teacher capacities for student assessment

According to the national curriculum for teacher education degrees in Basic (SEP, 2002) and Lower Secondary Education (SEP, 2010), teachers should be trained to perform student assessment in the classroom during their initial teacher education. Capacity to effectively 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 (SEP and INEE, forthcoming).

This course is aimed at preparing future teachers for organising teaching activities and assessing both progress and difficulties in student learning. Future teachers should learn how to use instruments adequate to curricular content and student characteristics. The course should also prepare them to assess learning processes, establish the timing for administering standardised instruments and write test questions requiring student reflection which use the knowledge and intellectual skills they have acquired. Future teachers are supposed to learn that the main objective of assessment is enhancing teaching and learning rather than just assigning marks (SEP, 2002, 2010; SEP and INEE, forthcoming).

Regarding in-service teacher training and professional development, classroom-based student assessment is a major line of work for the Directorate General of Continuous Training for In-service Teachers (DGFCMS) of the SEP. The main national initiative in this area, the annually published *National Catalogue of Continuous Training and*

Professional Betterment for Basic Education,[1] includes a large offer of over 1 000 in-service and postgraduate courses, from short courses (40-hour duration) to Master's and Doctoral programmes (SEP and INEE, forthcoming). These courses, which are delivered by higher education institutions throughout the country, are evaluated and accredited by the DGFCMS. 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 current 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, a large number of courses focused on curricular subjects include new approaches, techniques and instruments for classroom-based assessment (SEP, 2011b; SEP and INEE, forthcoming).

SEP has also implemented other support devices to facilitate the development of teachers' capacities for student assessment. This includes books for teachers, a special section on assessment and self-assessment in textbooks and a variety of materials available at SEP'S website, such as research articles on assessment, assessment indicators and links to other relevant websites about assessment (SEP and INEE, forthcoming).

External student assessment for selection purposes

IDANIS was created in 1989 with the objective of providing information for the selection of students entering lower secondary education. It is administered and processed by the DGEP-SEP only in those states requesting it. The test is composed of around 60 multiple-choice and fill-in-the-blanks questions, assessing students' basic abilities in three areas: communication, use of mathematics and abstract reasoning. Test items are very traditional and assess quite simple abilities. Results are used for assigning new students into lower secondary schools, especially for selective schools. For example, as explained in the SEP-DF's website, in the Federal District, students entering lower secondary education express three preferences of specific lower secondary schools during the enrolment process. IDANIS results are the main criterion to then assign students to individual schools.[2] Consequently, these results are quite important for students, as they determine their access to selective lower secondary schools (SEP and INEE, forthcoming; SEP-DF, 2011). As observed during the Review visit, in other states the selection process is typically carried out at the school level, by each lower secondary school having more candidates than available places.

EXANI I, created in 1994 and administered by CENEVAL, is undertaken on a voluntary basis and students or institutions must pay for taking it. As EXANI I results are widely used by upper secondary schools for student selection, it also has a strong influence on students and teachers in lower secondary education. EXANI I is currently composed of two instruments. The first and traditional one is a selection test that measures general intellectual abilities. The second is a recently introduced diagnosis test that assesses major subjects' content that should have been learned during lower secondary education and are relevant for the next level. Both instruments are norm-referenced (*i.e.* test takers are compared to each other). As for IDANIS, EXANI I results are used by upper secondary schools for selection purposes (Vidal, 2009; SEP and INEE, forthcoming).

IDANIS and EXANI I exert a strong influence on students and teachers, who devote time to prepare for them, as they determine students' chances to attend the school of their preference at the next level. Furthermore, a search for "IDANIS" or "EXANI I" on the Web leads to several sites offering paid courses for preparing students for these

examinations. On the CENEVAL website there is an announcement saying that guides and materials for preparing the tests are available for free, but the same materials are offered for pay on other sites.

External student assessment for formative purposes

ENLACE is a major assessment endeavour, covering more than 14 million students from every public and private school in Mexico, from third grade (primary education) to ninth grade (lower secondary education). Since 2008 it also includes students leaving upper secondary education in reading comprehension and mathematics. At the upper secondary level ENLACE is of a different nature, as it tests competencies rather than curriculum content, and involves a separate planning and logistics.[3]

Test administration is externally controlled by parents in each school. Teachers supervise the test administration for a student group different from their own (so the supervision of their own students is avoided). Tests include between 50 and 70 multiple-choice questions and are aligned to the curricular content. Using a three-parameter Item Response Theory model, results are estimated using horizontal scaling with reference to an average of 500 points corresponding to the average student results in the first application of ENLACE (in 2006). Students are also placed into four performance levels: "insufficient", "basic", "good", and "excellent". "Insufficient" means that students lack the necessary knowledge and skills to continue learning the subject at a proper pace. "Basic" level students are proficient in only a small part of the knowledge and skills assessed in a subject and school grade, but enough to continue learning satisfactorily. "Good" level students are proficient in most of the knowledge and skills assessed in a subject and grade. "Excellent" level students are proficient in all knowledge and skills assessed in a subject and school grade (SEP and INEE, forthcoming).

Results are presented with scores and percentages of students by performance level and are available on the SEP website. Students are given a code to look at their individual results, but during the Review visit the OECD Review Team perceived that many students do not find out about their results through the Internet. A study by Mendoza Trejo (2010) reveals that only 27% of parents of primary school students and 31% of parents of lower secondary school students find out about ENLACE results through the Internet (in a context where 52% of parents of basic school students do not find out about ENLACE results at all). This happens in spite of the substantial efforts by SEP which sends printed reports to parents with their children's results. Aggregate results by school are also available on line and can be consulted by the general public. Schools' directors and teachers have access to individual students' results by test component, and can learn the right and wrong answers for each student, permitting them to identify the topics which are more difficult for the students (SEP and INEE, forthcoming). Teachers also receive printed reports with their groups' results. Posters with school results compared to other schools are sent to the educational community and are to be posted at a visible place in schools.

Assessment can be "low stakes" or "high stakes" (as defined by Messick, 1999). ENLACE was initially presented as a "low-stakes" assessment, but it has progressively evolved into a "high stakes" scheme. This is a consequence of other uses of its results, the most relevant being the use of student results in the National Teacher Career Programme (PNCM), so part of teachers' salaries are tied to their students' scores in ENLACE (see analysis in Chapter 4). While ENLACE has been used in the context of the PNCM since its inception in 2006, its original design was based on the formative use of its results for

the improvement of teaching and learning in the classroom (see, for instance, Zúñiga Molina and Gaviria, 2010).

Originally the SEP suggested the following "low-stakes" uses for ENLACE results:

- Activities to involve parents in supporting their children's learning, such as workshops and reinforcement of learning at home;
- Creating materials for improving teaching and promoting the exchange of experiences and good practices between teachers and states;
- Development of continuous education offerings based on ENLACE results;
- Special initiatives to strengthen schools with low ENLACE results;
- Actions for strengthening educational management, such as establishing learning standards at the end of each education level; and
- Promoting programmes for improving education quality (SEP and INEE, forthcoming).

But ENLACE rapidly became a "high stakes" assessment with the publication of results at the school level, school rankings published in the media, monetary incentives for teachers based on their students' ENLACE scores and students with the highest scores receiving public recognition. These other uses of ENLACE are quite distinct from those initially intended. They may be leading schools, teachers and students to devote a large amount of time to practising ENLACE tests (Backhoff *et al*., 2008; Loureiro, 2009), a perception also conveyed to the OECD Review Team by a large number of stakeholders including in the schools visited.

During the Review Visit, the OECD Review Team witnessed the strong influence of ENLACE on schools' lives in a wide range of aspects. Examples of behaviours which were consistently mentioned by personnel interviewed in the seven schools visited (and which were corroborated by a large number of the stakeholder groups interviewed during the Review visit), include:

- In every school visited teachers and directors attached great importance to ENLACE results.
- In every school visited students heavily practised for ENLACE using the tests set in preceding years.
- Some schools seek to stimulate improvement in student motivation to achieve by participating in promotional programmes developed by education authorities such as "Let's go for 600 points" (organised by the Federal Administration for Educational Services in the Federal District, AFSEDF).
- Some teachers expressed the view that, as ENLACE has no direct consequences for them, students do not make a great effort when taking the test. As a result, as conveyed to the OECD Review Team in the schools visited, some teachers may seek to motivate students by indicating that ENLACE results contribute to their overall achievement data, which is not the case.
- Many teachers expressed concern about ENLACE not taking into account the different circumstances and contexts schools and students face.

- Some teachers recognise a positive effect of ENLACE, as "it exerts a little pressure on teachers to make an effort to improve students' achievement".
- ENLACE led teachers to have some particular concern about low achievement students. It also generated a special concern about reading comprehension.

As for students with special educational needs, there are a few adaptations for the test administration. ENLACE allows a relative to be in the classroom and help the student without providing her or him the answers. In some cases, depending on the specific type of student needs, the questions can be read to the student or help may be received in writing the responses on the response sheet (SEP and INEE, forthcoming).

In the case of Indigenous students, it is supposed that the tests' linguistic and cultural pertinence is assured through the accompanying work of the DGEI (SEP and INEE, forthcoming). The DGEI is part of ENLACE's Technical Council since 2008 and has worked with the DGEP in order to improve the pertinence and reliability of the test for Indigenous students. This work has included reviewing over 2 500 test items. However, during the visits to schools, teachers argued that many items include wording or situations that are unknown by rural and Indigenous students. In 2008 the teachers of a school based in the state of Chiapas filed a complaint against the SEP at the National Council for the Prevention of Discrimination (*Consejo Nacional para Prevenir la Discriminación*, CONAPRED) for linguistic discrimination in the application of ENLACE. The complaint received the support of representative organisations such as the National Congress for Indigenous and Intercultural Education. In 2011, the CONAPRED issued a resolution in favour of the teachers, stating that the SEP should introduce modifications to ENLACE to ensure that it is not culturally-biased against Indigenous students. The SEP accepted the resolution and is currently working with Indigenous education organisations with a view to adapting ENLACE to the needs of Indigenous students (CONAPRED, 2011).

Policy initiatives related to classroom-based student assessment

In the context of the Alliance for Quality in Education (ACE), a major curriculum reform effort is taking place in Mexico, the Comprehensive Reform of Basic Education (RIEB). A unified Study Plan, which articulates course programmes from pre-school to the end of basic education, is being piloted and progressively introduced in schools. The reform aims to improve the coherence of the whole system, focusing on providing students with a comprehensive education, so that they can acquire the skills needed for life (OECD, 2011a; SEP, 2011c) (see Chapters 1 and 2).

The new curriculum is oriented by an explicit profile to be achieved by every student leaving basic education. This profile is composed of ten main aspects which are expected to align all curricular efforts and to serve as a benchmark for evaluating the efficacy of the educational process (SEP, 2011c) (see Chapter 2).

The 2011 Study Plan establishes that teaching should be oriented towards competencies for life, so it should comprise more than simply explaining concepts. Every course is organised into five blocks and "expected learning outcomes" (*aprendizajes esperados*) are explicitly stated for each block. There has been an important effort for setting curricular standards to be used as a benchmark for external and classroom-based student assessment at the end of the main cycles of basic education: pre-school, third grade, sixth grade and lower secondary education. Curricular standards express what students should know and be able to do at the end of each cycle. There are standards for

the main formative areas: Spanish, English, mathematics, sciences, reading ability in Spanish and digital abilities. Standards have been set with technical advice from the University of London and are aligned with PISA performance level 3 in reading, mathematics and sciences, as a target to be reached by students leaving lower secondary education (SEP, 2011c; SEP, 2011d).

In early 2011, work on the design of a new generation of ENLACE assessments aligned with the new RIEB's standards and expected leaning outcomes was launched. These were piloted in 2011/12 and will be generalised to the full cohort of students as of 2012/13 when the RIEB is extended to all grades. This work has brought together the SEB, DGAIR, INEE, state educational authorities, teachers and experts under the co-ordination of experts of the Universidad Autónoma de Baja California.

Under this reform, a great effort is being made to establish a new approach to classroom-based assessment. More emphasis is placed on assessment as an on-going process and an essential part of teaching and learning. Teachers remain in charge of assessing students' achievement. Official documents explicitly establish that in basic education, the formative approach should have prevalence in all assessment activities, as the main objective is to improve students' performance. Teachers must explicitly and clearly explain to students and parents the expected learning outcomes for students and the assessment criteria, as well as the steps to be taken by students in order to overcome their difficulties. Teachers should not give a mark to students without advice on how to improve their performance (SEP, 2011d; SEP and INEE, forthcoming).

As for the instruments to be used for classroom assessment, the 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 (SEP, 2011c; SEP and INEE, forthcoming). Students should be frequently involved in self-assessment and peer assessment activities.

In the context of this new emphasis on classroom-based assessment, the SEP is working on a new approach to reporting marks to students and parents. The existing Report Card will be replaced by a Basic Education Card, which should combine qualitative and quantitative perspectives on student achievement, focusing on student progress in relation to the expected learning outcomes for each curricular block (SEP, 2011c; SEP, 2011d; SEP and INEE, forthcoming).

Marks will continue to be expressed using the current numeric scale ranging from five to ten, but each mark will have attached a description of the level of performance reached by the student, in relation to the expected learning outcomes. Performance levels will be labelled with an "A" for "outstanding" (corresponding to a 10); a "B" for "satisfactory" (corresponding to an 8 or a 9); a "C" for "sufficient" (corresponding to a 6 or a 7); and a "D" for "basic" or "elemental". For each mark the card includes a statement about the type of support needed for the student, which should be complemented with teachers' observations.[4] This new instrument has been developed by a working group formed by the SEP and the INEE in 2009. At this stage it is being trialled in 5 000 primary schools and 1 000 lower secondary schools.

Strengths

Teachers are committed to student learning

During visits to schools the OECD Review Team formed the view that most teachers in Mexico are genuinely involved in students' learning and achievement. In all meetings students consistently mentioned their teachers' determination in explaining the subject themes to them and in helping underachieving classmates. The OECD Review Team witnessed cases where teachers devoted time to extra classes and activities, as well as to meetings with parents, on a voluntary basis, on Saturdays or after hours. Teachers are also aware of the importance of adapting teaching and assessment to students' individual needs and cultural contexts – although the way they do it is not always adequate, as will be explained below. In some schools teachers asked advanced students to act as monitors in helping their classmates.

Teachers in Mexico play an important role in student assessment, as both formative continuous assessment and summative assessment are an essential part of their professional responsibilities. Assessment in Mexico is integral to the work of teachers. Evidence on student learning is collected regularly and a variety of aspects are taken into account for student assessment: tasks, effort, presentations, tests, projects. Nevertheless, there is much room for improvement, both in the way teachers approach formative feedback and in the way marks are established.

A new and comprehensive framework for classroom-based student assessment with an increased emphasis on outcomes is developing

The RIEB is bringing into the education scene a sound approach to classroom-based assessment. It constitutes a major and clever effort to align curriculum across compulsory education, from pre-school to the end of lower secondary education. While maintaining room and flexibility for adaptations to different educational levels, backgrounds and students' special needs, the RIEB organises study programmes with a comprehensive and coherent approach. The new curriculum seems to have a good balance between what is compulsory and what can be locally adapted.

One of the main features of the RIEB approach is the curricular alignment around students' expected performance. Each "block" for each course includes "expected learning outcomes", which are intended to be the main guide for teachers' work. Expected learning outcomes are aligned with curricular standards set at the end of each of three-year cycles (pre-school, first to third grade, fourth to sixth grade and seventh to ninth grade). Curricular standards are aligned with a graduate profile defined for students at the end of basic education (*i.e.* lower secondary education), which includes PISA's performance level 3 as a main component (SEP, 2011c).

By unifying curricular efforts around expected learning outcomes, the RIEB is generating a positive move from a content-based curriculum to a competencies-oriented one. This constitutes an important step forward for Mexico. Having PISA competencies in mind as part of the graduate student profile is clever in the sense that authorities are sending a unified message to teachers about what they are expected to generate as competencies.

The RIEB also includes a clear and interesting approach to student assessment in the classroom, for both formative and summative purposes, which is explicitly stated as a professional responsibility of teachers. As described previously, the RIEB: expands the

meaning of assessment, conceiving it as an essential part of teaching and learning; proposes the use of a wide range of assessment instruments; emphasises the formative purpose of classroom-based assessment; and, as will be explained below, introduces a critical shift in giving a new meaning to marks.

As observed by the OECD Review Team during visits to schools, teachers who have been in touch with the new curriculum – through direct participation in pilot experiences or through continuous training courses on the RIEB – are beginning to introduce new practices in their classrooms: checklists oriented to aspects such as participation, collaboration and research attitudes of students; rubrics; new authentic products being required from students (*e.g.* producing an announcement as a written task); projects oriented towards expected learning outcomes; and student self-assessment and peer assessment.

There is progress in aligning marks to expected learning outcomes

Another important shift that the RIEB is introducing in classroom-based assessment concerns the changes to the Report Card (renamed as "Basic Education Card"). The significant development is the intent to give a new meaning to marks in terms of expected learning outcomes.

The on-going approach is quite smart, in the sense that the focus is not on changing the scale, but on attaching a new meaning to the existing marks in terms of student performance. In some reforms of performance scales, a lot of energy is devoted to changing the scale, for example from numbers to letters. This is usually just a cosmetic change, and marking practices remain the same. The main problem with marking is the absence of a meaning for the different marks and the prevalence of marking as a matter of assigning points across a number of aspects and averaging them into a meaningless final mark (Wiggins, 1998; Ravela, 2009a). Current efforts in Mexico are being devoted to the establishment of a relationship between marks in the current scale and performance levels, although specific descriptions have not been developed yet. The new Basic Education Card is still being piloted and there is still work to do in order to define the specific meaning of each performance level for each of the subjects, so that marks acquire a clear meaning for teachers, students and families.

However, it must also be said that the new reporting scheme includes some debatable aspects such as indicators on reading speed and comprehension (a specific required classroom test consisting of measuring the number of words a student is able to read in one minute), and the practice of averaging marks for different subjects in lower secondary education.

There is good attention to reducing grade repetition in primary education

Grade repetition is a major problem in Latin America. Analysis for primary education in Latin America indicates that Guatemala (26%) and Colombia (21%) show the largest proportions of students in ages 7 to 11 who are lagging behind (*i.e.* not in the grade they should be at their age), with the figures for Brazil, Uruguay, Chile and Argentina being 11%, 10%, 9% and 6% respectively (OEI, 2010). The percentage for Mexico (3%) was the lowest within 17 countries.

It is well known that: *(i)* students who repeat a grade continue to be underachievers during their schooling (*i.e.* repeating a grade does not have the supposed effect of taking repeaters to the same level of their classmates); and *(ii)* students who repeat a grade

increase their probability of dropping out of school. Many countries have been trying to eliminate repetition, but in many cases the perverse effect has been that students move forward in their schooling without acquiring the expected learning (Torres, 1995; Schiefelbein and Wolff, 1992).

In Mexico, students cannot repeat between the first and the second grade of primary education (SEP, 2011a). In the context of the RIEB, this rule will be extended to the third grade. The emphasis will be on timely compensatory actions for underachieving students, instead of grade repetition. A remarkable fact is that, when looking at international student performance data (UNESCO's Latin American Laboratory for Assessment of the Quality of Education, LLECE), Mexico is above the regional average in reading and mathematics, both in third and sixth grades: 25% of third grade Mexican students are at or below level one in reading (below the regional average of 32%). Only Cuba, Costa Rica and Uruguay have a better performance than Mexico (UNESCO/OREALC, 2008). So evidence suggests that the Mexican approach to grade repetition is working well, as students are not lagging behind and, at the same time, they are achieving a good performance level in the regional Latin American context.

Mexico's grade repetition rates are higher in secondary education, where 22% of 15-year-old students have repeated at least one year during their schooling (OECD, 2010). However, with 22%, Mexico shows indeed the lowest percentage within Latin American countries participating in PISA. In countries like Argentina, Brazil, Colombia and Uruguay, around 35% of 15-year-old students have repeated at least one year (OECD, 2010). Mexico's figure is also lower than those for Spain and Portugal (both around 35%) (see Annex D).

The involvement of parents in their children's learning is being promoted

During its meetings and visits to schools, the OECD Review Team formed the view that school directors and teachers in Mexico are quite aware of the importance of frequent communication with parents. Although it is possible that in some cases parents are contacted in a merely administrative manner, in every school visited there was a concern about communicating marks to parents every two months and involving them in the support to low performance students. There were also noticeable initiatives to involve parents in current efforts for promoting reading and to encourage their active participation in the Councils for Social Participation in Education.

The current implementation of a new Basic School Card is also relevant, as it intends to give clearer meaning to marks in terms of students' expected learning outcomes. It is also positive that the new Card encourages parents' support in the learning of their children.

Some initiatives foster the development of teacher competencies for student assessment

As described earlier, there are some incipient but important initiatives to foster the development of teachers' competencies for classroom-based assessment. This area is increasingly present among the continuous teacher training offerings, although teachers are not yet satisfied with the quality of the courses offered. More recently, INEE and SEP started work on the development of instruments for classroom assessment which will be available on the Web. A toolbox for teachers is being developed by INEE and 1 million copies will be distributed to teachers. These initiatives are supposed to support the implementation of the RIEB and the new approach to assessment. Nevertheless, it is

important to note that it is not enough to distribute materials. The challenge is to make sure that teachers at schools have the opportunity to collectively interact with the new materials and with colleagues and try the new practices in their classrooms.

The capacity for designing and implementing large-scale assessments is remarkable

External student assessment is widely present in Mexican schools. For over 20 years IDANIS has been administered to students leaving primary schools and, through EXANI I, the same occurs with students leaving lower secondary education since 1994. During the 1990s, large-scale national standardised assessments were carried out in the context of the National Teacher Career Programme (PNCM). Also, EXCALE and ENLACE were launched in 2005 and 2006 respectively. Millions of standardised assessments are administered, processed and reported in Mexico every year.

Although the diversity of external student assessments is somewhat confusing for teachers, as will be explained later, they contribute towards focusing teachers' attention on students' achievement. Teachers look at each student's results and often focus on those with lower ENLACE results. Another effect has been teachers' awareness of the importance of reading comprehension for students' lives and the ability to answer test questions. Several teachers mentioned the importance for students to understand the text of the questions so that, for instance, they understand what is required in mathematics problems. Many teachers feel ENLACE results challenge their professional competence, in the sense of being pressured to improve their students' achievement (as measured by ENLACE).

IDANIS and EXANI I also play a role in motivating students and teachers around achievement. Nevertheless, there are some issues that must be addressed around the alignment between the different external assessments and the new curriculum.

Each of the external assessments involves much technical work through a range of committees in charge of developing instruments and processing data. There is also a large logistical capacity installed. This capacity exists not only within SEP and INEE but also within CENEVAL, which has vast experience in organising standardised assessments. The experience with external assessment has relied almost exclusively on multiple-choice tests leaving room for the introduction of more complex types of tasks which could send important new pedagogical signals to teachers, students, parents and schools.

Challenges

There is a need to change the culture of teaching

Improving the way teachers assess their students involves a critical change in established teaching practices. As stated by Stigler and Hiebert (2009), teaching practices are difficult to change because they are part of a culture. Teachers learn how to teach mainly through informal participation in school and classroom practices over long periods of time – and less so during initial or continuous training programmes. Teaching is a practice that is learned by living in a culture more than through formal studies. The way teachers do things in schools is determined rather by cultural scripts which are like the DNA of teaching. For these reasons changing teachers' practices is a complex and long-term endeavour (Stigler and Hiebert, 2009). Keeping this in mind is crucial for Mexico's efforts to change assessment practices and make curriculum changes effective. It is not just a matter of writing a new study plan, improving the Report Card, offering new

continuous training courses or delivering new materials to schools. Instead, a major cultural change must be promoted.

The Mexican educational system is highly centralised. Schools and teachers are the subject of control by supervisors, the state and federal level authorities, as well as the teacher union. Simulation is a strong part of the system's culture, as the more you are controlled, the more you need to simulate what you are doing to meet the expectations. During the Review visit, the word "simulation" was mentioned frequently as being part of the culture by a variety of stakeholders: authorities, supervisors, school directors, teachers and researchers. This fact should be carefully taken into account by those who are leading the changes at the central level. In such a culture it is even more difficult to introduce real changes in classrooms by decree.

Teaching to the test is prevalent across the school system

Teaching to the test has become a widespread pedagogical practice in Mexico. School directors, teachers and students consider that practising standardised tests is the best strategy for improving student achievement. This was consistently conveyed by personnel interviewed in all the seven schools visited by the Review Team and confirmed by many of the stakeholder groups interviewed during the Review visit. Schools motivate students for improving their ENLACE results, including through the participation in formal programmes such as "Let's go for 600 points" (organised by the Federal Administration for Educational Services in the Federal District, AFSEDF). Several weeks before the administration of ENLACE, teachers devote considerable class time to what they call the "Pre-ENLACE", which are testing sessions similar to the actual test using examples from previous years. Something similar happens with IDANIS and EXANI I. In some schools, teachers and students devote significant amounts of time to the practice of these tests. While ENLACE results have consequences for teachers and directors – part of their salaries depends on them (see Chapter 4) – IDANIS and EXANI I are important for students and parents as the possibility to attend a selective school depends on their results. Similarly, the SEP also strongly encourages schools to prepare students for PISA tests. The publication "Towards PISA 2012" (*Hacia PISA 2012*) is widely distributed in a version for teachers (250 000 copies) and another for students (4 million copies), encouraging secondary school students and teachers to practise weekly with PISA released items during the whole school year. Obtaining good results in PISA is presented to students as a matter of national pride and loyalty to Mexico (SEP, 2011e and 2011f).[5]

The critical issue that should be carefully considered is whether the objective is simply to raise average scores (in either ENLACE or PISA) or, instead, improve Mexican children's wide range of competencies. It is very important to make a distinction between the objective and the indicator. As Linn and Gronlund (2000) put it:

> *We are almost always interested in making inferences that go beyond the specific test. We would like, for example, to be able to say something about the degree of understanding of mathematical concepts based on the score that is obtained on a math concepts test. Because the items on a test only sample the domain of interest, the test score and the inference about the degree of understanding are not the same. A generalisation is required, and it is the generalisation, not the test score per se, that is important. When the specific items on the test are taught, the validity of the inference about the student's level of achievement is threatened.*

The nationally promoted strategy of encouraging students to practise ENLACE and PISA items may lead to an increase of scores in the short term, but it is not clear that it will lead to sustained better learning for the student population in the longer term. Regrettably, the risk is that teaching-to-the-test practices reinforce the cultural belief that practising standardised tests is the path to improving learning. This is made worse by the incentives teachers and school directors have to promote teaching to the test as ENLACE results have a direct impact on their salaries (as explained in greater detail in Chapter 4). These incentives are not necessarily aligned with the best strategies to sustain the process of improving students' knowledge, abilities and attitudes. Important educational objectives, which are not assessed in the tests, are neglected. As standardised tests cover a limited range of competencies and cross-curricular skills, teaching to the test narrows students' learning experiences (see Morris, 2011, and Rosenkvist, 2010, for a discussion). It would be unfortunate that teaching-to-the-test practices undermine the many positive effects ENLACE has had in the Mexican education system such as the greater focus on improving student outcomes, the greater attention to students with learning difficulties and the transparency of student results for education stakeholders (as explained earlier in this chapter and in Chapter 2).

There is an excessive reliance on multiple-choice tests

A major problem in external student assessment in Mexico is the almost exclusive use of multiple-choice tests, with potential distortionary effects on the education of children. The extent to which Mexican external student assessment relies on multiple-choice questions should be a concern for education authorities, because this can narrow students' vision about what counts as performance. The success of the RIEB requires the introduction of more complex tasks to external assessments as well as a greater variety of assessments.

Similarly, IDANIS and EXANI I should introduce greater variety to the type of questions asked, so they can assess the competencies promoted by the RIEB. The argument that these tests assess "abilities" instead of curricular content is quite outdated. If Mexican authorities are to maintain these instruments for selection purposes (which, in itself, is debatable) or just as an effective administrative tool to distribute students across lower secondary schools (as seems to be the case in the Federal District), they should align these assessments to the RIEB's expected learning outcomes and standards, in order to reinforce a unified message to teachers and students about what is important in learning.

ENLACE has too many objectives

As explained earlier, while ENLACE was originally supposed to be a diagnostic and formative assessment instrument, new objectives and consequences were added subsequently, the most important of which is the use of its results to provide monetary incentives to teachers and school directors (see Chapter 4 for more detail). As explained by Linn (2000), assessment systems that are useful for formative and monitoring purposes usually lose much of their credibility when high stakes are attached to them, because the unintended negative effects of the high stakes often prevail over the intended positive effects.

During the Review visit, teacher, school management, students and educational experts raised instances of some of the non-desirable effects of ENLACE. Examples include time diverted from regular curriculum for special test preparation for ENLACE;

practising test items without analysing them in depth; and difficulties in ensuring the integrity of test administration. As documented in Zúñiga Molina and Gaviria (2010) and OECD (2011b), experiments conducted in Mexico suggest that test cheating may occur at a significant level - between about 4 and 10% of overall percentages of probable test cheating cases for grades 3 to 6 in the period 2006 to 2009. As Zúñiga Molina and Gaviria (2010) put it "Each year, the ENLACE test is becoming more important in terms of social and media impact, and this might be reflected in the increase of cheating behaviours." In their analysis of the opportunities for the further development of ENLACE for evaluation and teacher incentives in Mexico, Zúñiga Molina and Gaviria (2010) conclude:

> *… since its initial implementation, the [ENLACE] program results have been used for different purposes, despite the repeated warnings that appear in official documents of the program regarding the need to avoid some of those uses. In that sense, it can be argued that some of the most visible uses of the program are related to purposes that have been considered as inadequate by those who have been in charge of design and operation of the program. Such is the case, for example, of the frequent use of ENLACE, in the media, to establish national or state-wide rankings of schools, based exclusively on the average scores achieved by students at each particular school; or the occasional use that education officers, and other concerned parties, have made of the tests results as if they were an unequivocal indicator of the quality of work carried out by teachers.*

Also, even if EXCALE is the student assessment specifically designed to monitor student learning objectives at a system level (over time and across states) (see Chapter 6 for further details), this function is *de facto* also being accomplished by ENLACE as a consequence of the use education authorities and the media make of the results. This is not desirable as ENLACE not only assesses students on much more limited curricular content than EXCALE but it also uses a smaller number of items. As stated by one of the stakeholders interviewed by the OECD Review Team, "Census kills sample". ENLACE has become much more visible than EXCALE and is being used at all levels with a large variety of objectives (formative and diagnostic role with students, system evaluation, school accountability, state accountability, monetary incentives for teachers and school directors).

As explained earlier, ENLACE has brought considerable benefits such as further teacher concentration on student achievement, particularly that of underperforming students, or greater awareness of the importance of reading comprehension. But unintended effects of ENLACE seem to be significant. In spite of the large amounts of data collected, the extent to which those data are being used formatively is not clear. This calls for an important reflection about the uses ENLACE results should have in order for the multiple benefits of ENLACE not to be undermined.

Teachers have a narrow approach to teaching and formative assessment

Although teachers are aware of the importance of the formative dimension to classroom-based assessment, the OECD Review Team found little evidence of it being implemented in an adequate manner. Teachers seem to have a narrow understanding of formative assessment. Giving feedback to students is conceived as giving them marks or points for a task, telling students whether their work was acceptable or not, or asking them to revise their work or make extra effort (Ravela, 2009a). All of these are concepts profoundly embedded in teachers' culture. Formative assessment basically consists in

giving students a general indication about what was wrong with the test or the task assessed. As stated by students in interviews with the OECD Review Team, "they explain to us what we did wrong", and "they tell us what we need to improve".

Formative assessment is essentially a matter of appropriate feedback. It should permeate the process of teaching and learning instead of being something that happens after learning. In the same sense, feedback and its use should not be something that occurs between assessments. It should be construed at the core of what is being assessed. Feedback should be continuous and immediate (Wiggins, 1998). During the school visits the OECD Review Team formed the view that in Mexican teachers' culture there is lack of reflection on these issues. There is little awareness of the importance of giving feedback during the learning process and little knowledge of cognitive learning processes. For example, on analysing test results teachers focus on the items instead of the cognitive processes involved. There was also little evidence of students reflecting about their own processes of learning, as well as of authentic tasks (Wiggins, 1998). As stated in a recently published study, there is the need in Mexico for teachers to give students more descriptive feedback (García *et al.*, 2011, pp. 33-34).

Student self-assessment and peer assessment practices are also at an early stage of development (García *et al.*, 2011). Teachers rarely show students samples of good work so they can both understand the performance level expected of them and assess their own work. Also, teachers do not typically show samples of weak pieces of work which progressively were transformed into good quality work (García *et al.*, 2011, pp. 75-76).

Supervisors, directors and technical pedagogical advisors (ATPs) also have a narrow approach to classroom observation. Their main concern is to control the administrative aspects of teaching. These include teachers' punctuality; control of students' attendance; remembering the topic of the previous class; writing on the blackboard the topic for the present class; controlling discipline in the classroom; facilitating students' participation in classroom activities; checking students' homework; and managing content for the 50 minutes of class. The more specific aspects of teaching and students' learning processes receive less attention.

Marking practices lack pedagogical significance

As described earlier, current marking practices are deeply ingrained in the teaching culture. Marking in Mexico consists of assigning points to students across a range of tasks and behaviours and then averaging them to obtain a mark. Points for students' work are assigned in a normative way, by comparing students within classes and giving the greatest number of points to the "best" student, regardless of whether or not the standards are indeed excellent.

Marks assigned by averaging points across a range of tasks and behaviours usually lead to a kind of *grand* number with no clear significance (Ravela, 2009a). Students "earn" points for attending classes, doing homework, participating in teamwork, giving a presentation and taking an exam. Points are assigned by the teacher in a quite subjective way, as there are no rubrics specifying the meaning of points for each of these activities. As a result, a "7" may mean quite different things in different schools, in different classrooms within a school and, even worse, for different students within the same classroom and for the same student in different moments of the school year. Each individual teacher determines the marking criteria but it is not guaranteed that the same teacher will be consistent in the application of the criteria across students and over time. Also, as there is not a clear statement describing the kind of performance to be achieved,

students cannot understand what is expected from them. So obtaining a high mark becomes the main objective for them, possibly distorting education efforts (Shepard, 2006). Parents are also more worried about the marks than about their children's real learning.

An issue which deserves special attention is assigning marks by comparing students within the classroom, in absence of explicit standards or benchmarks, which leads teachers to adapt their expectations, their cognitive requirements and their teaching to the current "level" of the students in their classroom. The result is teaching less to the more disadvantaged (the issue of adaptations to different populations and cultural contexts will be further elaborated below).

The practice of combining the assessment of effort and motivation with the assessment of actual achievement is also an issue to be addressed, because it undermines academic marks as indicators of performance (Ravela, 2009a; García *et al.*, 2011). This practice also leads to students simulating effort, as they quickly learn how to behave with each teacher in order to make a good impression. Finally, marks become a disciplinary instrument for teachers, instead of a tool to inform about learning. Controlling students' behaviour through marks is not the same as creating a motivating learning environment (Shepard, 2006).

All these features of marking are strongly and profoundly embedded in teachers' culture. As with every cultural practice, it will be challenging to change marking practices. A central standardised resolution for the whole country is likely not to work. Schools and teachers must have room to do their work and try new approaches to marking within the national framework. Discussions are needed at the local and school levels about the different processes involved: discipline, attendance, homework, testing, the kind of tasks and so on. As said earlier, it is also crucial to develop a more sophisticated view on cognitive processes. Every actor at every level should be involved: students, parents, teachers, directors, ATPs, supervisors and heads of sector. Otherwise, the existing cultural norms for marking will persist.

Instruments for reporting marks need further improvement

The instruments used for recording and reporting students' achievement are of critical importance in every educational system and at every level. They play several functions: communicating what is expected as learning to students, teachers and parents; motivating; and giving direction to students' efforts.

As said earlier some important changes in the report cards are being promoted by the SEP in the context of the RIEB. However, there are still a number of aspects that need to be addressed. First, as stated above, there is the need for more specific and detailed descriptions of student performance at the different levels. It is crucial to ensure that parents and students can understand the information provided by the report card.

Second, the new Basic Education Card's approach to stating observations and orientation for students and parents seems to be too standardised. There is the need to conceive a more flexible instrument, with more room for teachers to communicate what they expect and how the student may improve.

Third, the new Basic Education Card includes some aspects of concern, particularly the standards for reading speed. The rationale for giving this indicator so much visibility is not clear. During meetings with the OECD Review Team a number of students expressed their anxiety and pointed to the meaninglessness of the measure: "To read fast

is more difficult as I can't understand what I am reading". The way reading comprehension features in the Basic Education Card is also of concern. Performance is considered adequate if the student answered correctly three questions out of four.[6] However, nothing is said about the type of text, its length and complexity, or about the difficulty of the questions to be answered. There is a clear need for a more sophisticated approach to reading assessment.

Finally, it should be noted that for lower secondary education the practice of averaging marks for different subjects and units, which has no real meaning, is kept as central to the marking scheme. Also, averaging bimonthly marks into a unique final mark does not recognise students' progress over the year (García *et al.*, 2011).

Student assessment leads to little interaction among teachers

Teachers are rarely brought together around student assessment issues. Teachers typically do not interact in the preparation of assessment instruments or the development of marking criteria. In fact, the exchange of classroom practices is quite unusual among Mexican teachers: in schools teachers work in relative isolation from each other, even if less so in primary schools. Visiting each other in classrooms is rather uncommon. According to the TALIS survey only 27.5% of teachers in Mexico participate in a professional development network, being the type of activity least mentioned as a professional development activity (see Annex D).[7] It seems that only in small communities, namely small multi-grade schools, do teachers systematically share materials, experiences and assessment instruments.

Most teachers do not exchange their assessment instruments with others, do not discuss expected learning outcomes and do not develop a shared approach to marking. This seems to be another cultural feature of teaching in Mexico which might become an obstacle for the RIEB efforts, including the implementation of a new approach to classroom assessment. Moreover, teachers in Mexico seem to be little aware of their need to learn more about student assessment. When asked about the areas in which they have a "high level of need" for professional development, only 15% mentioned "student assessment practices" and 13.7% mentioned "content and performance standards" (see Annex D).

Student assessment is not being used as a professional development activity for teachers or a way for them to improve their professional judgment. Moderation which involves authentic student work is underdeveloped and should be a key strategy for teachers training, as well as for building a new culture around classroom assessment. Moderation of assessment and marking also has the potential for establishing links to classroom practices.

There is a lack of consistency of student assessment across schools and classes

Schools have no explicit marking criteria and typically do not have documentation on their approaches to student assessment. This fact, together with the absence of moderation procedures for aligning the meaning of teachers' marks, leads to a situation in which the meaning of marks differs from one region to another, from one school to the next and even from one classroom to another classroom within the same school and from one student to another within the same classroom. In the words of one student interviewed by the OECD Review Team: "The meaning of a 10 varies from one teacher to another and it depends on showing your effort".

Relationships between marks assigned by teachers and ENLACE results have not been analysed. Studies of this kind would lead to a picture of the correspondence between classroom-based marks and results in external student assessments, as well as to an idea about variability in the meaning of marks for teachers.

There are limited capacities at the state and local levels to support classroom-based assessment

An impressive effort is being undertaken at the central level to promote the RIEB, including continuous teacher training, a novel approach to marking, a toolbox for classroom assessment, and adjustments to external student assessment and textbooks' content. But Mexico is a large and diverse country. As teachers' competencies for formative assessment and marking are limited, there is the need for supporting strategies on the part of state and local authorities. It will not be enough to deliver materials to schools around the country. State authorities, heads of sector, supervisors, heads of teaching must play a role.

However, at the state level there seem to be limited capacities for assisting teachers in pedagogical and assessment issues. Supervisors' work is mainly focused on checking compliance with school regulations. There is a clear need to change the profile of supervisors and ATPs, towards a more specialised role in pedagogical issues and assessment (see also Chapter 5). At the state level, instead of replicating the functions of INEE, there is the need to develop agencies to provide direct support to individual schools, involving an interaction around pedagogical, didactic and assessment issues. This could also include strategies for schools and teachers to collaborate and exchange experiences (see Chapter 5). Alongside this, the SEP should consider strategies to improve capacities at the state level to provide much more meaningful support to schools on those aspects which bear a greater promise to improve student learning.

Making assessment inclusive for students remains a challenge

Basic schools in Mexico have an important proportion of students with special needs in mainstream classes with little extra support. Both ENLACE and EXCALE do not have a developed strategy for including this population. Test implementation and application variations based on educational services or adaptations for students with special educational needs are not in place, except in some particular instances for ENLACE (SEP and INEE, forthcoming) (see also Chapter 6 concerning EXCALE).

Also, there is an important concern among teachers and other stakeholders about the cultural bias of external student assessment. The OECD Review Team heard several references to the use of items which include situations or terms which are totally unfamiliar to students who do not live in urban settings or who have an Indigenous background. As explained earlier, this problem was formally recognised through the resolution by the National Council for the Prevention of Discrimination (CONAPRED) stating that ENLACE is culturally-biased against Indigenous students and that further work by the SEP is needed to make it a fairer assessment for those students.

The RIEB approach to diversity is adequate, as it establishes common standards and expected learning outcomes but, at the same time, gives room for local adaptations according to students' backgrounds, needs and context. Nevertheless, there are risks in the way teachers interpret and apply the necessary adaptations. As observed by the OECD Review Team during school visits, many teachers working with disadvantaged students do not interpret the need to adapt the curriculum in a pedagogical sense. Instead, their

adaptations for those students consist of lowering their expectations of achievement and of reducing the cognitive challenge (for example, asking those students to read shorter texts or to use less complicated numbers in math calculations). Other teachers interpret the need to make adaptations as adapting their marking criteria, giving disadvantaged students a better mark than the one they would obtain with the set of criteria established for the whole class. Hence there is an important risk around the interpretation and application of adaptations, involving lower expectations for disadvantaged students.

Policy recommendations

Ensure a coherent and comprehensive strategy for the RIEB implementation

In the present context the improvement of classroom-based assessment in Mexico needs to be developed alongside the implementation of the RIEB, given the inclusion of relevant initiatives related to both formative and summative assessment. Additionally, the implementation and impact of the RIEB crucially depend on the successful introduction of changes in student assessment practices and on aligning these with the expected learning outcomes and standards defined in the new curriculum.

The implementation of a new curriculum is a long-term endeavour that should be carefully and cleverly designed. During the past two decades Latin America has had several experiences with ambitious curricular reforms with little impact in classrooms. The enterprise of ensuring that the reform reaches the classroom cannot be left solely to the central authorities, especially in a country as large and diverse as Mexico.

Particular attention should be given to ensuring that the breadth of the curriculum and learning goals established in the new Study Plan is maintained in student assessment by making sure that all subject areas and objectives are given certain forms of attention. This involves not only classroom-based assessment, but also external assessments (see below). As for classroom-based assessment, teachers need to integrate in their practices a much broader range of activities and instruments, to promote and capture more complex cognitive processes.

Consolidate teachers' command of learning and formative assessment

The successful implementation of the RIEB requires the introduction in Mexican teaching culture of a more refined vision of learning processes and didactic issues, a precondition for introducing new formative assessment practices. Teachers should receive support and training to move from a rather traditional view of teaching, conceived as explaining themes and concepts, towards a broader concept based on the facilitation of learning and the development of competencies. In this context, the repertoire of approaches to learning and assessment needs to be expanded, moving away from assigning lots of exercises and practising tests.

Formative assessment is intrinsic to good teaching practices. Any effort to improve teaching and learning in Mexico must involve the improvement of teachers' competencies in formative assessment in a thoughtful and consistent way. Currently most teachers use three types of approaches to what they call formative assessment: *(i)* indicating to students their mistakes in a test or a task; *(ii)* asking students to make more effort; and *(iii)* giving students praise in order to motivate them (Picaroni, 2009; Ravela, 2009a).

But authentic formative feedback involves other things. As Wiggins (1998) puts it,

> *Safeguarding the core premise that assessment should improve performance, not just audit it, requires that assessment embody and demand self-adjustment based on good feedback... the moment when the student understands why some part of his or her work is a mistake is entirely different from the moment when the student perceives that the teacher does not like that part of the work. The best feedback is highly specific, directly revealing or highly descriptive of what actually resulted, clear to the performer, and available or offered in terms of specific targets and standards.*

Formative assessment also involves the development of instruments such as rubrics to make students reflect by themselves about the gap between what they were expected to achieve and their actual performance (Wiggins, 1998; Ravela, 2009a).

It is important to recognise that expanding teachers' repertoire of practices and instruments for formative assessment is not just a matter of sending new materials to schools or loading them onto a website (and expecting teachers to use them). To ensure successful implementation it is crucial to create networks of teachers and develop sustained professional interaction around the new assessment procedures (Ravela, 2010). Designing and introducing good instruments and practices for formative assessment requires collaborative, continuous and interactive work.

Develop a new approach to marking

Taking responsibility for the certification of students' achievement is inherent to the professionalism of teachers. Parents and society trust teachers' accurate and comprehensive judgment about students' achievement. External assessments may contribute to this function, but cannot replace teachers' professional judgment.

In Mexico, marking criteria are defined by individual teachers and are not documented. As a result, the meaning of marks is quite unique to each school, classroom and student. If student marking is to be aligned with the RIEB's expected learning outcomes and standards in a consistent way across the country, then a priority is to establish mechanisms for the moderation of marking, both within and across schools. The objective is to reduce the variations in the ways teachers assess students and set marks so that equity of student assessment is improved. Moderation strategies should include frequent interaction between teachers around the meaning of marks (within and across schools), focused on the relationship between marks and performance levels for each block in the curricular content, as well as on the kind of appropriate evidence for each of the performance levels. Moderation of marks may also include statistical analysis of correlations between teacher-based marks and student results in external tests such as ENLACE, IDANIS or EXANI I. However, moderation should not be understood as a way of standardising marking. There must be room for flexibility and locally-based decisions in assigning marks. The important issue is that the meaning of the relationships between expected learning outcomes and marks is clear and shared.

Moderation has the potential to provide a very powerful professional learning opportunity for teachers that they can relate closely to their classroom practices. Moderation also contributes to improving teachers' professional judgments about student work and their developing a shared understanding of marking criteria or standards. Evidence of the powerful benefits of professional discussions around students' work to improving students' learning outcomes has been demonstrated in New Zealand's

programmes of professional development in literacy, numeracy, and assessment for learning (Timperley *et al.*, 2007).

The development of moderation processes should go along with the development of guidelines at the national level for assessing against student learning objectives. Teachers require exemplars of student work to illustrate achievement at different levels or marks, benchmarks or indicators of desired student achievement, optional assessment tasks, and tests.

Other relevant issues that need to be addressed are:

- The procedures to come to a decision on the mark for a particular student (other than averaging points);
- Getting away from normative approaches to marking (*i.e.* comparing students within a class) and understanding the meaning and importance of criterion-based assessment (*i.e.* giving a mark against established standards for the different performance levels); and
- Recognising the importance of reporting separately on student achievement and attitudes/engagement.

Distinct aspects of students' performance should provide for separate assessment reporting, so what needs to be improved is clearer. For example, it is important to distinguish between the ability to understand what is being asked in a mathematics problem, the ability to do calculations or attitudes in the classroom.

Working around real problems and challenges faced by teachers when marking students may be a powerful training strategy to develop teachers' capacities. Workshops in which teachers independently analyse and mark samples of students' work and then compare and discuss the marks assigned by each one of them is a good example of the kind of training needed. As will be explained below, the implementation of these strategies requires a large number of specialists in assessment at the local level, continuously visiting schools and working with directors and teachers.

The ability of teachers to mark against national student learning objectives should also be assessed in the context of school evaluation. This could involve comparing teacher-based marks to ENLACE results, reviewing the instruments and criteria used for marking, assessing the extent to which marks are related to levels of performance in the expected learning outcomes, and examining whether marks are clearly communicated to students and parents.

Develop a sound strategy for strengthening teachers' capacities for student assessment

Developing teachers' skills and competencies for student assessment requires a major investment and wise planning. As stated by Crozier (1989), "investment in human resources is the most difficult to do. But in case of success, it is the most effective of all".

Continuous professional development should be conceived as much more than teachers individually taking in-service courses. The approach to professional development should also involve interaction between teachers within schools and across schools at the local level, and be highly focused on teaching practices. According to TALIS, participating in "courses and workshops" is the most common type of professional development activity undertaken by teachers in Mexico (see Annex D). At

the same time, participation in networks is the least common professional development activity (see Annex D).

Similarly, improving student assessment skills during initial teacher education requires more than delivering courses on assessment. The core strategy for improving future teachers' skills in assessment should be the implementation of a whole new approach to assessment in teacher education institutions. During their studies future teachers should be assessed, both for formative and summative purposes, in the same way they will be expected to assess their students: using rubrics, complex tasks, marking criteria based on performance levels, explicit expected learning outcomes, and so on. This requires considerable investment in teacher education programmes on how to give feedback and how to assign marks in a criterion-referenced approach.

The same is valid for SEP's current efforts to provide support for teachers through the publication of documents and articles about classroom-based assessment practices, and suggested classroom assessment practices through the SEP website. While these are important initiatives, they are not enough to promote real change. This kind of strategy, if isolated, is like "shooting to the sky and waiting for a duck to fall" (Ravela, 2010).

Also, in order to improve teaching practices, there is a need to move away from the conception of teaching as an isolated activity towards a vision of teaching as a professional activity, involving interaction with colleagues and open to peer review (Ravela, 2011). Teachers should be expected to reflect on their own practice and learn from experience. At the core of a teacher development strategy there should be a space for teachers to experiment, share and reflect on their classroom practices (Shepard, 2006; Ravela, 2009b).

A significant effort is needed in training teachers in the development of assessment rubrics and other kind of qualitative instruments for assessing students' daily work in a more meaningful way, as well as in approaches to marking learning units. An important aspect is that it is not sufficient to produce instruments at the central level and send them to the schools. Teachers themselves must be involved in producing their own instruments, within the new approach to assessment.

Redesign and strengthen the role of supervisors

The role of states and their supervision system is crucial for effecting change at the classroom level given their proximity to schools. States take responsibility for education services within their boundaries and supervisors are the main link between schools, authorities and educational policies. The success of any national reform crucially depends on the capacity at the supervision and state levels to ensure the necessary links to classrooms. In this context, there is a need to invest substantially in the capacity of supervisors, ATPs, heads of teaching and heads of sector so they can substantially contribute to the implementation of reforms. An area of particular focus should be instructional and pedagogical leadership, including sound strategies for classroom observations.

The visits to the schools by the OECD Review Team revealed two major features of the work of supervisors, as asserted by interviewed stakeholders. First, supervisors are crucial to the implementation of change, given their authority over and proximity with schools. Second, supervisors are not necessarily reliable in effecting pedagogical and instructional change given their focus on the political and administrative control of schools. The OECD Review Team formed the view that the main concerns of supervisors

seem to be related to collecting information and being in control of the activities within schools. Their pedagogical interventions are rudimentary (see also Chapter 5).

At the same time, in a system which lacks the tradition of teachers' collective work around pedagogical issues and teaching practices, there is an enormous need for some specific agents to facilitate and lead teachers' interaction (Ravela, 2009b). These agents should bring teachers together at the school or local levels, promote and lead the exchange between teachers, build a common framework on assessment, create spaces for collaboration, frequently observe classrooms and give teachers external feedback on their own practices. Given the present reality, the agents within the supervision structure (supervisors, ATPs, heads of teaching, heads of sector) seem to be ideally placed to become such agents. This would involve redefining their role so it concentrates much more on instructional and pedagogical leadership (see also Chapter 5). This could also include creating new positions and recruiting new people with adequate training.

Some states, such as Aguascalientes, are already developing experiences in this area. In this state, supervisors are taking responsibility for creating networks of schools and building alignment across the different levels. Another ten states are now adopting similar practices (OECD, 2011a). These correspond to current efforts to strengthen support to schools: "the reorganisation of the educational system in Mexico relies in large part on the creation of Regions for the Management of Basic Education (RGEB) which are geographical units defined around the school to support various aspects related to educational services such as planning, programme implementation, resource distribution, data collection, distribution of materials, assessment and accountability. Each region (RGEB) will have an Educational Development Centre (CEDE) charged with ensuring that the local administrative and academic conditions are appropriate to support improvement in school performance and student learning outcomes" (OECD, 2011a). It is crucial that the CEDEs have a clear emphasis on promoting school and teacher networks around pedagogical and assessment approaches, practices and instruments. It is important to bear in mind that there is an important risk that the existing culture absorbs the innovation potential of CEDEs and these become a new administrative centre for the control of schools.

Promote the formative use of standardised student assessments

A policy priority should be to promote the adequate formative use of standardised student assessments such as ENLACE and PISA, including getting away from the incentives given to schools to practise the tests. For example, a document entitled *Suggestions for the pedagogical use of ENLACE results* has been widely distributed. Its main objective is to use ENLACE results as a pedagogical-technical tool for teachers to improve their teaching practices and help enhance the quality of classroom learning (SEP 2011g; SEP and INEE, forthcoming).[8] However, the extent of the use of this document by teachers is unclear to the OECD Review Team. Whenever asked about the way they tried to improve ENLACE results, the answer was most often "by practising the tests". Also, the fact that results of ENLACE become available to students, parents and teachers only the school year following the application of ENLACE does not facilitate the formative use of results (Mendoza Trejo, 2010). In these circumstances, ENLACE results are less relevant to inform strategies to improve the learning of individual students. Hence, there should be a reflection about improving the timeliness of results' delivery so they can inform learning strategies in the same school year ENLACE is taken.

In the case of PISA, authorities should focus teachers' attention on understanding its framework – what PISA assesses – and on reflecting and discussing how to develop the assessed competencies in the classroom. External assessment uses for improving teaching should not be focused on test items but instead on the assessment conceptual framework. The main point is that teachers should understand what is being assessed and why, so that they reflect on ways to improve their teaching practices (Ravela, 2010, 2011).

Develop a more articulated and coherent framework for external assessment

While the RIEB includes a sound framework for classroom-based assessment, external assessments are quite diverse. IDANIS, EXANI I, ENLACE and EXCALE are not clearly articulated within a strategy for external assessment. Each of them emerges in a different period, in response to different historical needs, but there has not been an effort to clearly redefine and articulate their role within the student assessment framework.

It is recommended that the following issues are addressed:

- The purposes of ENLACE should be revised and clearly communicated. ENLACE is currently used for a great variety of purposes, including the monitoring of the system at the national and state levels, a task more appropriately achieved by EXCALE.
- All the external assessments should be redesigned and aligned with the RIEB's standards and expected learning outcomes, and be oriented towards competencies. In the case of ENLACE, there is some progress in this respect and a new generation of ENLACE assessments will be introduced in 2012/13 following their piloting in 2011/12. In the case of IDANIS and EXANI I, their focus on "abilities" that "predict" future performance seems to be somewhat outdated. If these exams for selection purposes are to be kept – which, in itself, should be an issue for consideration – their content should also be aligned to the RIEB.
- EXCALE will need to develop a more complex design, in order to both maintain the achievement trends initiated in 2005 with the test aligned to the old curriculum and assess students' achievement in the new competencies fostered by the RIEB (see also Chapter 6). It should be noted that this work has now started with the use of subsamples in 3rd grade of lower secondary education, as of 2012.
- A significant effort should be undertaken to introduce more diverse types of tasks in external assessments, not just multiple-choice questions. This includes not only the external student assessments mentioned above, but also assessments administered for teacher appraisal (see Chapter 4 for further details). If a new overall approach to assessment is to be adopted by teachers, the omnipresence of multiple-choice questions should be reduced. In teacher appraisal tests, teachers should also have the experience of being assessed with instruments and tasks that go beyond multiple-choice questions, so that they replicate that in their own assessment of students.

Although introducing constructed response items or other kind of complex tasks in large-scale assessments is quite demanding, technology today makes things more affordable. In Chile, for example, an important capacity has been developed around electronically codifying open-ended questions and students' written responses in the national standardised test (System to Evaluate the Quality of Education, SIMCE), as well as in codifying teachers' portfolios in the teacher appraisal programme (Santiago *et al.*,

forthcoming; Manzi *et al.*, 2011). While Mexico has already developed a significant capacity in designing, administering and processing multiple-choice tests, the next challenge should be to develop capacity to introduce constructed response items and more complex tasks in large-scale assessments. Making this important shift to constructed response items and complex tasks in external assessments, together with the effort to align them with the RIEB's focus on competencies, should be priorities for the short term.

Develop strategies to address the detrimental effects of ENLACE

As described earlier, ENLACE is a dominant element of Mexico's education system. It has brought considerable benefits to student learning in Mexico but it has also generated considerable unintended effects. As a result, a major priority for policy should be the development of strategies to eliminate, or at the very least reduce, the current detrimental effects of ENLACE. This effort should be informed by an in-depth study of the impact of ENLACE on practices in schools and classrooms.

One strategy could be reducing the high stakes of ENLACE. A range of options are possible to achieve this depending on the extent to which stakes for school agents are reduced. A possibility is to rethink the objectives of ENLACE, including a return to the original motivation of ENLACE as a purely diagnostic and formative tool for student assessment. Another possibility is to add to this original objective some role in system evaluation to assess whether, at the national level, student learning objectives in the subjects covered by ENLACE are achieved or not. Most OECD countries limit the use of standardised student assessments to these two functions. If the objective of using ENLACE for school accountability (publication of ENLACE results at the school level) and teacher appraisal is maintained, then it is imperative to develop value-added techniques to capture the real impact of individual schools (for the publication of results and their use in RNAME) (see also Chapter 5) and considerably reduce the weight of ENLACE results in teacher appraisal for the reasons explained in this chapter and Chapter 4. Alongside this, it is important to monitor the potential unintended effects of the high-stakes uses of ENLACE through appropriate research studies.

Another strategy is to transform ENLACE into a tool for the external summative assessment of students, *i.e.* an external examination system. This would involve extending the range of student learning objectives assessed by including more subjects and broadening the range of tasks assessed. It would also have consequences for students, as with the contribution to final marks or as a certification mechanism at the end of key stages in education (such as end of educational cycles). This would introduce a strong motivation for students (Shepard, 2006; Messick, 1999) and also for teachers, because most teachers are genuinely concerned with their students' success. As stated in a recent OECD report on lessons from countries with high performance in PISA, high-stakes assessments for students at the end of certain levels of the educational system introduce strong incentives for students and teachers. And this reinforces what is being called horizontal accountability – teachers being held accountable to their colleagues and to parents – instead of vertical accountability – teachers responding to authorities and administrative instances (OECD, 2011c). Horizontal accountability is less easily simulated than vertical accountability, because of the daily face-to-face relationships.

An approach to reduce the burden for schools, teachers and students as well as the costs of administration would be to administer the assessment only at key stages of education, such as at the end of each educational cycle (*i.e.* 3rd, 6th and 9th grades), instead of every single grade as is currently the case. Also, another possibility is to leave at least

part of the marking to teachers alongside sophisticated moderation procedures. This could involve exchanges of teachers between schools, centrally designed rubrics and marking manuals, the central control of the assessment administration and the central marking of a sample of schools. The experience of Sweden (Nusche *et al.*, 2011) with its external student testing and of New Zealand (Nusche *et al.*, 2012) with its certification system at the upper secondary level can be particularly useful.

This approach may have several advantages as it would:

- Rely further on teachers' professionalism and promote it;
- Consist of a valuable professional learning experience for teachers;
- Involve teachers in national-level assessment and foster their understanding of the RIEB;
- Promote student effort;
- Make possible the use of a wider range of assessment tasks and questions; and
- Allow the assessment of a broader set of subjects and of a broader set of competencies within subjects.

Ensure student assessment is inclusive

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 and the format and content of external standardised tests (such as ENLACE and EXCALE) should be sensitive to particular groups of students such as Indigenous students, students with special needs, and students living in disadvantaged social contexts. In the context of Mexico, this is indeed a formidable task given, for instance, the existence of 68 Indigenous languages.

The cultural background should be carefully taken into account in test design, to prevent the use of words, expressions and situations which are unfamiliar or completely unknown in certain cultural settings. It is suggested that quality assurance guidelines are prepared and practices adopted that ensure that external assessments are evaluated or reviewed for their potential bias in these respects. This may include consideration of a variety of assessment formats (test-based, performance tasks, oral, written) so that individual students/groups of students are not systematically disadvantaged; and peer review of the content of test questions. Also, there is a need to provide for special adaptations for students with special needs to take the tests.

Finally, regarding teacher-based assessment there is a dilemma around marking criteria and local adaptations of curriculum and expected learning outcomes. On the one hand, every child should achieve the stipulated expected learning outcomes for each grade and subject. This reflects the objective that teachers have the same academic expectations of children regardless of their socio-economic background. But, on the other hand, large cultural and socio-economic differences are part of a complex and unequal Mexican society. So there is the need for closely working with teachers on reaching the

right balance between not excluding students from learning (as a result of too demanding expectations) and not lowering expectations for their learning.

Improve reporting to students and parents

The commendable introduction of the new Basic Education Card needs to be accompanied by some adjustments so that it becomes an authentic instrument for learning. An initial adjustment is to make statements associated with performance levels more specific. At the moment, these are quite general. For example, for "C" or 6-7 (in the old marking scale), the statement is "shows a sufficient performance in the expected learning outcomes for this block", regardless of the school grade, subject or learning block. It would be beneficial for this general statement to be grade- subject- and block-specific, in association with the concrete learning outcomes students should achieve.

Expected learning outcomes for each grade, subject and block should be clearly explained to students and parents, including with examples of what is an acceptable, a satisfactory and an outstanding performance. A version of the expected learning outcomes for each learning block should be developed for parents, so they can understand the meaning, for example, of "shows a sufficient performance in the expected learning outcomes for this block". Chilean curricular maps of progress may be an interesting inspiration for this endeavour (Santiago *et al.*, forthcoming).

Another key issue is the need for more detailed information about individual student performance (García *et al.*, 2011, p. 90). The space dedicated to it in the new Basic Education Card might prove not that useful if teachers are not prepared to give students and parents more precise indications about how to improve the student's performance. The risk is that this instrument may remain limited to teachers' call for greater student effort. Valuable examples of approaches with greater potential to generate student progress are the experiences of Denmark and Sweden with individual student plans containing detailed and specific recommendations for each student (Shewbridge *et al.*, 2011; Nusche *et al.*, 2011).

Notes

1. The National Catalogue of Continuous Training and Professional Betterment of Basic Education In-Service Teachers 2011-2012 is available at *http://formacioncontinua.sep.gob.mx*.
2. Information about the process of assigning students to schools in the Federal District is available at *www.sepdf.gob.mx/principal/archivos/nota_preg_frecuentes_ingreso_secundaria_2011.pdf*.
3. However, it should be noted that the analysis in this report refers only to ENLACE at both the primary and lower secondary levels.
4. Basic Education Cards are available at *www.boleta.sep.gob.mx*.
5. Both publications are available at *www.pisa.sep.gob.mx/descargas.html*.
6. Basic Education Cards are available at *www.boleta.sep.gob.mx*/.
7. OECD's Teaching and Learning International Survey was implemented in 2007/08, covering lower secondary education and with the participation of 23 countries (OECD, 2009). The results derived from TALIS are based on self-reports from teachers and directors and therefore represent their opinions, perceptions, beliefs and their accounts of their activities. Further information is available at *www.oecd.org/edu/talis*. TALIS results for Mexico are provided in Annex D.
8. ENLACE publications for each grade are available at *www.enlace.sep.gob.mx/ba/apoyos_para_el_uso_pedagogico*.

References

Backhoff, E., E. Andrade, A. Sánchez and M. Peon (2008), *Estudio Comparativo del Aprendizaje en Sexto de Primaria 2005-2007,* INEE, Mexico, *www.inee.edu.mx/index.php?option=com_content&view=article&id=3469&Itemid=1051.*

CONAPRED (Consejo Nacional para Prevenir la Discriminación) (2011), Resolución por Disposición 1/2011, EXP. CONAPRED/DGAQR/ 419/08/DR/II/CHIS/R289, *www.cneii.org/images/stories/resolucion-por-disposicion.pdf.*

Crozier, M. (1989), *Estado Modesto, Estado Moderno. Estrategia para el Cambio*, Fondo de Cultura Económica, Mexico.

DOF (Official Federation Gazette) (1994), *Agreement number 200, which Sets Standards for Student Assessment in Primary, Lower Secondary and Normal Schools*, 19 September 1994, SEP, Mexico, *www.sep.gob.mx/work/models/sep1/Resource/3f848841-8476-4596-9835-d71e306221a3/a200.pdf.*

DOF (2009), *Agreement number 499, which Amends Agreement number 200, which Establishes Student Assessment Norms in Primary, Secondary and Normal Education*, 4 November 2009, SEP, Mexico, *http://dof.gob.mx/nota_detalle.php?codigo=5117051&fecha=04/11/2009.*

EPPI (Evidence for Policy and Practice Information and Co-ordinating Centre) (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, Evidence for Policy and Practice Information and Co-ordinating Centre, London.

García, B., J. Mejía and A. Meza (2009), *Propuesta para Evaluar y Reportar el Proceso de Desarrollo de Competencias de los Alumnos de Educación Básica, mediante una Nueva Boleta de Calificaciones*, SEP, Mexico.

García, A.M., M.A. Aguilera, M.A. Pérez and G. Muñoz (2011), *Evaluación de los Aprendizajes en el Aula – Opiniones y Prácticas de Docentes de Primaria en México*, INEE, Mexico.

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

Linn, R. and N. Gronlund (2000), *Measurement and Assessment in Teaching*, Prentice Hall, New Jersey.

Loureiro, G. (2009), *Evaluación en el Aula, Currículo y Evaluaciones Externas*, Partnership for Educational Revitalization in the Americas (PREAL), Santiago de Chile, *www.ucu.edu.uy/Default.aspx?tabid=1267.*

Manzi, J., R. Gonzalez and Y. Sun (eds.) (2011), *La Evaluación Docente en Chile*, MIDE-UC, Santiago de Chile.

Mendoza Trejo, F. (2010), *Estudio Sobre el Impacto de la Difusión de los Resultados de la Prueba ENLACE 2009*, Instituto de Fomento e Investigación Educativa, A.C.

Messick, S. (1999), "Key Issues", in F. Ottobre (ed.), *The Role of Measurement and Evaluation in Education Policy*, UNESCO, Educational Studies and Documents N° 69, Paris.

Morris, A. (2011), "Student Standardised Testing: Current Practices in OECD Countries and a Literature Review", *OECD Education Working Papers*, No. 65, OECD Publishing, *www.oecd.org/edu/workingpapers*.

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

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

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, *www.oecd.org/edu/talis*.

OECD (2010), *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices*, Volume IV, OECD Publishing, *http://dx.doi.org/10.1787/9789264091559-en*.

OECD (2011a), *Progress with Educational Reform in Basic Education in Mexico: An OECD Perspective*, OECD Publishing.

OECD (2011b), *Establishing a Framework for Evaluation and Teacher Incentives: Considerations for Mexico*, OECD Publishing.

OECD (2011c), *Strong Performers and Successful Reformers in Education. Lessons from PISA for the United States*, OECD Publishing.

OEI (Organización de Estados Iberoamericanos) (2010), *Metas Educativas 2021 – La Educación que queremos para la Generación de los Bicentenarios*, OEI, Madrid.

Picaroni, B. (2009), *La Evaluación en las Aulas de Primaria: Usos Formativos, Calificaciones y Comunicación con los Padres*, Partnership for Educational Revitalization in the Americas (PREAL), Santiago de Chile, *www.ucu.edu.uy/Default.aspx?tabid=1267*.

Ravela, P. (2009a), "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*.

Ravela, P. (2009b), "La Evaluación del Desempeño Docente para el Desarrollo de las Competencias Profesionales", in E. Martín and F. Martínez Rizo, *Avances y Desafíos en la Evaluación Educativa*, OEI-Fundación Santillana, Madrid, pp. 113-126.

Ravela, P. (2010), *¿Qué Pueden Aportar las Evaluaciones Estandarizadas a la Evaluación en el Aula?*, Document No. 47, Partnership for Educational Revitalization in the Americas (PREAL), Santiago de Chile, *www.ucu.edu.uy/Default.aspx?tabid=1267*.

Ravela, P. (2011), *¿Qué Hacer con los Resultados de PISA en América Latina?*, Document No. 58, Partnership for Educational Revitalization in the Americas (PREAL), Santiago de Chile, *www.ucu.edu.uy/Default.aspx?tabid=1267*.

Rosenkvist, M. (2010), "Using Student Test Results for Accountability and Improvement: A Literature Review", *OECD Education Working Papers*, No. 54, OECD Publishing, *www.oecd.org/edu/workingpapers.*

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

Schiefelbein, E. and L. Wolff (1992), *Repetition and Inadequate Achievement in Latin America's Primary Schools: A Review of Magnitudes, Causes, Relationships and Strategies*, UNESCO/OREALC, Santiago de Chile.

SEP (Secretariat of Public Education) (2002), *Licenciatura en Educación Primaria. Plan de Estudios 1997. Documentos Básicos*, SEP, Undersecretariat of Basic and Normal Education, Directorate General of Standards, 4th printing, Mexico, *www.dgespe.sep.gob.mx/sites/default/files/planes/lepri/plan.pdf.*

SEP (2010), *Licenciatura en Educación Secundaria. Plan de Estudios 1999. Documentos Básicos*, SEP, Undersecretariat of Basic and Normal Education, Directorate General of Standards, 4th printing, Mexico, *www.dgespe.sep.gob.mx/sites/default/files/planes/les/plan.pdf.*

SEP (2011a), *Normas de Control Escolar Relativas a la Inscripción, Reinscripción, Acreditación, Regularización y Certificación en la Educación Básica*, SEP, Directorate General of Accreditation, Incorporation and Revalidation (DGAIR), Mexico, *http://dgece.sev.gob.mx/docs/NORMAS_BASICA_2011-2012.pdf*.

SEP (2011b), *Catálogo Nacional. Formación Continua y Superación Profesional para Maestros de Educación Básica en Servicio 2011-2012*, SEP, Directorate General of Continuous Training for In-Service Teachers (DGFCMS), Mexico, *http://formacioncontinua.sep.gob.mx/index.php?option=com_content&view=article&id=234:catalogonaciona20112012&catid=35:novedades&Itemid=110*.

SEP (2011c), *Plan de Estudios 2011. Educación Básica*, SEP, Mexico.

SEP (2011d), *Acuerdo 592 por el que se Establece la Articulación de la Educación Básica*, SEP, Mexico.

SEP (2011e), *Competencias para el México que Queremos: Hacia PISA 2012. Manual de Alumnos*, SEP, Mexico.

SEP (2011f), *Competencias para el México que Queremos: Hacia PISA 2012. Manual de Maestros*, SEP, Mexico.

SEP (2011g), *Características Generales e Información de los Reactivos Aplicados para su Uso Pedagógico. Sexto Grado de Primaria 2011*, SEP, Mexico.

SEP (Secretaría de Educación Pública) and INEE (Instituto Nacional para la Evaluación de la Educación) (forthcoming), *Country Background Report for Mexico*, prepared for the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes, to be available from *www.oecd.org/edu/evaluationpolicy*.

SEP-DF (Secretariat of Public Education in Federal District) (2011), *Preguntas Frecuentes Ingreso a Secundarias 2011-2012*, SEP-DF, Mexico, *www.sepdf.gob.mx/principal/archivos/nota_preg_frecuentes_ingreso_secundaria_2011.pdf*.

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

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark*, OECD Publishing.

Stigler, J. and J. Hiebert (2009), *The Teaching Gap*, Free Press, New York.

Timperley, H., A. Wilson, H. Barrar and I. Fung (2007), *The Teacher Professional Learning and Development Best Evidence Synthesis*, Ministry of Education, New Zealand, *www.educationcounts.govt.nz/publications/series/2515*.

Torres, R. (1995), "Repetition: A Major Obstacle to Education for All", in *Education News, No. 15*, UNICEF, New York.

UNESCO/OREALC, 2008, *Los Aprendizajes de los Estudiantes de América Latina y el Caribe. Primer Reporte de los Resultados del Segundo Estudio Regional Comparativo y Explicativo*, UNESCO/OREALC, Santiago de Chile.

Vidal, R. (2009), *¿Enlace, Exani Excale o PISA?*, CENEVAL, Mexico.

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

Zúñiga Molina, L. and J. L. Gaviria (2010), *Challenges and Opportunities for the Further Development of the ENLACE Assessment for Evaluation and Teacher Incentives in Mexico*, expert paper commissioned by the OECD for the Co-operation Agreement between the OECD and the government of Mexico.