

Chapter 7

Assessment Reforms Around the World

Rita Berry

7.1 Introduction

In contrast to the pursuit of evidence at the end of the learning process, which largely defined the twentieth century approach of assessment, the international agenda for assessment in the twenty-first century shows signs of growing recognition of using assessment for learning purposes. There has been widespread call for new ways to think about assessment since high-stakes tests without supportive environments can harm learning (e.g. Black, 1998; Stiggins, 2004; Wiliam, 2006; see Chapter 11 by Scott). The calling has produced varied responses, ranging from a total abolition of high stake testing in some education systems to attempts to strike a balance between classroom and large-scale assessment in a synergistic system. Common to all these visions is the notion of assessment as a positive tool for learning and an interconnected part of teaching and learning. It is a pedagogy that is readily integrated into instructional designs (Berry, 2008). Over the last few decades, there have been waves of assessment reforms around the world. This chapter presents the assessment reforms in different educational contexts in different parts of the world. Selected cases will be presented to illuminate the issues brought to public attention in the reforms with a focus on assessment policies and practices. It examines the tensions and outcomes of assessment reform arising at the interface of policy and implementation and presents the experiences of some countries that turned the challenges into better teaching and learning opportunities.

7.2 The Changing Assessment Landscape in Europe, Americas and Australiasia

In the last half a century, Europe saw a number of education reforms that placed assessment reforms as an important issue on the reform agenda. In *Sweden*, for

R. Berry (✉)

Department of Curriculum and Instruction, Hong Kong Institute of Education, Tai Po, Hong Kong
e-mail: rsyberry@ied.edu.hk

example, the first wave of assessment reform began in the 1960s when there was a widespread belief that learning was something which could be quantified and measured. As a result, a norm-referenced grading system was introduced. Over the years of implementation, people constantly raised the question as to how much these grades could actually provide information about learning. With the view of knowledge and learning gradually migrating from positivistic and quantitative to hermeneutic and qualitative, the curriculum had become less focused on detailed knowledge and facts and more on constructs such as critical thinking, cooperation and problem solving. This resulted in the norm-referenced grading system being replaced by a goal-oriented, criterion-referenced one. Four grades were introduced to indicate progression of learning (IG – fail; G – pass; VG – pass with distinction; MVG – pass with special distinction) (Wikström, 2006). The idea is that the students should continue their education until at least a G grade has been reached and that the grade outcome should carry a formative function in addition to its designated summative use. Since the introduction of the criterion-referenced grade system, tests for scale calibration (the National Tests) have been available for the teachers to identify standards so that grades could be comparable. Still, teachers differed in scoring the tests as they had different interpretation of the rubrics (Nyström, 2004).

France initiated the “Haby” reform in 1975 with the goal of identifying and developing students’ true talents (Brauns & Steinmann, 1999). Notable among these initiatives was the virtual abolition of all public examinations below the 18+ Baccalauréat level (the final school leaving examination) together with the regular promotion tests during the course of schooling and their replacement with continuous assessment by the teachers (Broadfoot, 1985). French teachers assess their own pupils informally on a regular basis through oral or written exercises in the classroom or through homework. There is formal assessment in the higher forms but the teachers are given free rein on the frequency of the assessment and how they are marked (Bonnet, 1997). The purpose of assessment is to use the information obtained to adapt teaching to the needs of the students. However, the judgments on on-going work are made on the same basis as summative judgments. There is little written feedback of a formative nature (Raveaud, 2004). A large number of the teachers feel the pressure brought about by the implementation of continuous assessment. They are neither committed to, nor prepared for, these responsibilities (Broadfoot, 1985). Given that high-stakes public examinations remained in place for school leavers, students and teachers generally prefer to work to the examinations with teaching and learning more focused on conventional types of knowledge and competence (Bonnet, 1997).

For a long time, *Germany* used a national 6-point marking system (grade 1–6, where 1 is the highest) to monitor students’ achievements. Around 1960s, a strong critique of grades emerged because several empirical studies demonstrated that this form of assessment was not helpful for student learning (Ingenkamp, 1971). In addition, during this time, there was a shift in perceptions about learning that are commonly and internationally labelled as the need for “lifelong learning” and “learning-to-learn”. Education reformers called for the abolition of grades and for

the use of formative assessment. Consequently, several alternative tools for student assessment were proposed, all of which had a more formative focus, for example, in 1970, the Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States of Germany (Kultusministerkonferenz, KMK) decided that marks should be substituted by verbal reports in elementary schools, at least in grades 1 and 2. This decision was intended to base assessment on individual progress instead of social comparisons. Empirical studies of the implementation and practice of verbal reports in elementary schools, however, showed that the reform was not working as hoped. Valtin (2002) and Wagner and Valtin (2003) analyzed the effects of different types of assessment (marks versus verbal reports) on the development of educational outcomes in elementary school. The research comprised 241 children from East and West Berlin who were tested several times, individually or in groups, from grade 2 to grade 4. The outcomes were about attitude toward learning and toward school subjects, academic self-concept, achievement motivation, test anxiety, intelligence, and academic achievement in mathematics and German. Contrary to researchers' predictions, students did not profit notably from verbal reports. One reason for these findings, the researchers reported, might be that the teachers only practiced formative assessment when writing the reports but not in everyday situations in the classroom.

Before the enactment of education reforms between 1981 and 1986, assessment in *Greece* had been very summative-oriented and used mainly for accountability and selection reasons. The assessment approaches varied from end of term to final examinations, using numerical or grading as the main methods of recording and reporting. The overarching aim of the education reform was to make a change to the then traditional pedagogy to a more progressive child-centred one (Ministry of Education 1985). The educational reform agenda included the abolition of formal assessments, examinations and grading and unobstructed promotion from level to level. Mavrommatis (1996) conducted a study to investigate the implementation of assessment in Greek classroom. Twenty teachers were observed and then interviewed to obtain a general picture of the assessment practice the teachers used to assess their students. To enhance understanding of teachers' assessment practices, 360 serving and prospective teachers were invited to complete a questionnaire. The study revealed a number of difficulties that constrained Greek teachers from a full implementation of the assessment reform initiatives. In the Greek classroom, comparisons between students were often found to be an underlying classroom goal although official guidelines advised teachers to avoid this. A few teachers involved in the study did try to use assessment to help individual students learn better. These teachers made an effort to help students see what their learning gaps were and to make them aware what could be done to close the gaps. However, the teachers said that they could only do this occasionally because the constraints of large class size and the time taken up in dealing with many other teaching duties. Other issues revealed by this study included teachers' feedback use and assessment criteria. It was found that feedback was too general and short and therefore most of the students could not work out what kinds of actions they needed to take to improve. The teachers found it hard to achieve a clear understanding of their students' progress as

there was a lack of specific written reference criteria reflecting the national standards of prescribed objectives.

Believing that the school system developed for the period of dictatorship (1939–1975) was no longer appropriate for being a democratic member of the European Union, *Spain* initiated an education reforms in 1990 which held formative assessment at their heart. The first initiative in the reforms related to assessment included the abolition of the certification at the end of basic education. There is now only one state examination (*Selectividad*) which serves as the gateway to university education. Other times, assessment is classroom and teacher based. To investigate whether the formative assessment policy made an impact on teachers' assessment practice, Remesal (2007) interviewed fifty Spanish teachers. The results showed that there was a mismatch between the reform intentions and teachers' conceptions of assessment. The teachers, in particularly secondary school teachers, inclined strongly to associate assessment with accountability instead of linking assessment with teaching and learning.

As with Spain, *Portugal* saw the need to revamp its education system after the period of dictatorship. In 1986, the Assembly of the Republic of Portugal approved a four tier education system composing of (i) pre-school education (3–5 years old); (ii) basic education (6–14 years old); (iii) secondary education (15–17 years old); (iv) higher education (18 years old and above). From 1992 onwards, the Portugal government made it explicit in its legislation that formative assessment should prevail in the classroom at all grade levels, with the purpose of improving learning and teaching. According to the legislation, formative assessment should be an integral part of teaching and learning and be related to: (a) self-assessment and self-regulation of learning on the part of pupils; (b) the use of a diverse number of strategies and assessment instruments; (c) the participation of pupils and other intervening persons in the assessment process; (d) the transparency of procedures; (e) the definition of the criteria relative to developing competencies; and (f) the feedback that teachers should provide to their pupils in a systematic way (Fernandes, 2009a). However, Fernandes (2009b) found out in his study that formative assessment was yet to become a norm in teachers' classroom practices. Although most teachers in his study acknowledged the significance of formative assessment in student learning, they were in fact keener on designing tests simulating to those used in the external summative assessments.

Similar challenges have been identified in other countries in Europe and Americas. In *England*, Black and Wiliam (2005) point out that teachers' judgments do feed into national assessments, at 7, 11, 14 and 16, but concerns for reliability and accountability mean that such judgments are made in a way that has little impact on learning (see Chapter 2 by James). The government of *Netherlands* made schools accountable for student learning though it was met by widespread resentment from the teachers. Towards the end of the twentieth century, there was a growing pressure from the Dutch educational officials on schools to implement classroom assessment schemes based on norm-referenced tests. The purpose of the schemes was to systematically chart student learning progress over time. As the tests were standardized, it would be easier for the government to monitor school

performance by comparing students' scores across schools. In *Russia*, the main purpose of the assessment reform is to use assessment as a means to promote national standards. The government was determined to prepare the students for the rapidly changing socio-economic conditions in Russia. In 2003, the government introduced a national system of student assessment in the final year of secondary schooling in Russia which aimed at setting minimal standards and providing the much needed credibility to nationally recognized certification. *Denmark* increasingly believes that students need more testing to excel. They think the undesirable results in the international comparisons resulted from a weak assessment culture. The government subsequently set up the Danish Evaluation Institute and is considering establishing a central specification of learning targets with a new marking scale (Egelund, 2005). In the *United States*, multiple demands for accountability lead the country into measuring the amount of learning that has taken place, which provides little insight into how it might be improved. The American vision of long-term stability as a value and a goal associated with education – an evolutionary not revolutionary approach to educational reform appears to have been interrupted by the urgency surrounding the demands of the “No Child Left Behind” Act of 2001 and its mandated thirst for large-scale assessment (Hess & Petrilli, 2006) (see Chapter 3 by Flaitz). Similar situation happened in Latin America where *Brazil* and *Chile* also used assessment as mechanisms to monitor education systems (Carrasco & Torrecilla, 2009; Guimarães de Castro, 2001).

Although the above-mentioned countries undertook different initiatives in their assessment reforms, most of them shared one commonality – advocating the use of assessment for learning. With all these good intentions, the results of the reforms showed that there were tensions between government assessment policies and classroom assessment practices. Teachers were still inclined very strongly to associate assessment with accountability instead of linking assessment with teaching and learning.

Some countries achieved better outcomes in their assessment reforms. In 1968, *Finland* underwent an education reform with continuous assessment being used at the basic school level for guidance and encouragement purposes and on student learning and growth (Frassinelli, 2006). All assessment of student learning is based on teacher-made tests, rather than standardized external tests. The teachers viewed regularly scheduled teacher-made classroom tests as opportunities for learning as much as for assessing student achievement. Grades are prohibited by law and only descriptive assessments and feedback are employed (Sahlberg, 2009). The non-grade approach is to encourage students to become responsible, make their own decisions, and learn to plan their own life (Aho, Pitkänen & Sahlberg, 2006). In recent years, the focus of reform has been on the need for new type of life-long professional training for teachers to include up-to-date research, virtual learning environments and changes in the work force. It is worth noting that Finland relates the success also to their dedicated teachers who are willing to continuously strive for professionalism. Finland related its excellent student results of the Programme for International Student Assessment (PISA) in 2000 and 2003 to the success of its national school reform.

Canada advocated striking a balance between large-scale testing and classroom assessment and to use both to facilitate student learning. Common features among jurisdictions in the Canadian Report prepared by the Council of Ministers of Education in 2005 include:

- providing tools teachers need to develop and implement a well-planned student evaluation program that uses assessment techniques for formative, diagnostic and summative purposes;
- developing achievement standards for subject and grade specific courses that are supported by formative and summative assessment tools;
- promoting alternative approaches to student assessment and the education of educational personnel to adopt and effectively utilize such practices in the classroom;
- providing rubrics and exemplars to teachers as guides to varying levels of student performance;
- developing provincial processes regarding the assessment of learners;
- providing sample assessment strategies for classroom use;
- providing teacher professional development opportunities for all teachers; and
- promoting the use of criterion-referenced evaluation as a means of classroom-based evaluations.
- using the results of large scale assessments in a formative manner to guide academic intervention initiatives and to improve student learning.

(Council of Ministers of Education, 2005)

Beginning in the 1990s, in Canada, province-wide assessment systems were in place in most provinces for measuring and reporting on student achievement in literacy and mathematics at the school, school district and provincial levels (Dunleavy, 2007). In the classroom, the government advised that assessment should make up a large part of the school day, not in the form of separate tests, but as a seamless part of the learning process (Friesen, 2009). An important key to shifting the classroom, school, or district to a stronger learning orientation is to focus professional learning towards a passionate interest in helping learners become more self-regulated, more motivated, and more successful, which many schools across Canada were engaged in helping learners achieve this goal (Kaser & Halber, 2008).

New Zealand, influenced by local and overseas developments, in particular from the United Kingdom (see Chapter 2 by James) and to a lesser extent Australia (see Chapter 5 by Klenowski), implemented its major curriculum and assessment reforms affecting primary and secondary schools in 1989 (Philips, 2000). From this time, school curricula have been extensively restructured listing achievements objectives by levels (1–8). Criterion-reference (more commonly called “standards-based assessment” locally) was introduced to replace norm-referenced assessment. The main rationale for these changes was to improve student learning through better designed and more focused teaching and assessment programmes. The programmes were seen as helping teachers as they provide them with a more structured system for guiding teaching and monitoring students’ learning progress. With the

encouragement from the Ministry of Education and the Education Review Office, teachers and schools tried to come to grips with the new system. To implement the assessment initiatives successfully, Crooks (2002) drew teachers and researchers' attentions to some details in assessing students.

- The teacher's judgement might be made on the basis of just one task, yet many tasks could be developed for the objective and students would perform differently on different tasks.
- Children who could do a particular task on 1 day often could not do that task or a very similar one the next day.
- Trying to complete this process for all the achievements objectives in the primary school curriculum for a particular class was overwhelmingly time consuming and threatened the quality of teaching.
- There were major difficulties in summarizing student performance by aggregating across achievement objectives in a curriculum strand or whole curriculum areas, with student performance fluctuating markedly across objectives.
- Teachers differed considerably in the standards they set for judging what an objective had been met or a level achieved.
- The gap between adjacent levels (2 years of normal progress) was too large to give a satisfying sense of progress (pp. 243–244).

7.3 The Assessment Reform Experiences in Asia and Africa

Asia has a long tradition of using examinations to select government officials and to assign people of different talents to different professions. On record, China was the first country that used scholastic achievement tests as a means to select its civil servants (Han & Yang, 2001). From Western Zhou, the first dynasty in China over 2,000 years ago, to Qing Dynasty, the last dynasty in Chinese history, imperial examinations were used frequently for selection purposes (Berry, 2008). The imperial examination system had a far reaching impact on its neighbours, as countries such as Vietnam, Korea and Japan established their own imperial examination system based on the ideas borrowed from China (Wang, 2008). In *Vietnam*, beginning in the eleventh century, the examinations were conducted personally by successive kings who pursued Confucian ideals (Broadfoot, 2009). As with the countries in the western world, Asian countries underwent educational reforms with new policies set for assessing their students. The reforms in *Mainland China*, *Hong Kong* and *Taiwan* aim at making a change to the examination-oriented education to an education that is aimed at all-round development in students. Teachers are encouraged to use assessment to enhance teaching and learning. However, the findings of a number of studies revealed that there were gaps between intentions and reality. In many classrooms, teaching was still very examination-driven (see Chapter 4 by Berry).

South Korea experienced a widespread expansion of education between 1945 and 1970, when the government decided to establish a national education system that aimed at providing educational opportunities to all school aged children and high

quality human resources to the society. The system, highly centralized, is responsible for developing national level standardized tests and diagnostic tests for basic skills of elementary students. The college entrance examination is extremely high-stakes. Most South Korean students spend their entire high school life preparing for this examination. Fierce competition amongst students was overtly encouraged. To achieve good results, students attend privately owned institutions after school. Statistics showed that seven out of ten students receive private tutorial for an average of 6.8 h a week and private expenditure for education accounts for an average 12.7% of household expenses (Na, 2005). In the international comparative tests, South Korean students outperformed many of their counterparts from the resource affluent countries. Given the amount of stress that the students face, the price of success is quite high. South Korean high school students suffer from high rates of depression and suicide cases particularly around times of major examinations.

In *Japan*, the secondary school and university entrance examinations exert considerable influence on assessment practices in the classroom. To prepare students for the examinations, Japanese school teachers have traditionally relied heavily on summative assessment of student learning. Standardized paper-and-pencil tests are the most common form of assessment used in the school. Assessment has been and remains dominated by teacher-centred practices (White, 2009). There were some individual attempts to make assessment serve teaching and learning. Yoshinori and some of his colleagues used extended assessment tasks to facilitate deep thinking. In the process, the educators became aware of what their students needed and used the information to improve teaching (Shimizu & Lambdin, 1997). The major assessment reform agenda in Japan was in higher education in the 1990s with “Outcomes Assessment” as the main reform focus. Universities were required to constantly check their activities and enhance the quality of education by themselves (Kiamura, 1997). It was a response to a twofold interpretation of assessment needs realized in Japan about a decade ago. The interpretation tried to address two issues – “accountability” and “student active learning”. Japanese universities had been described as “hard to enter, easy to graduate from” and it was deemed necessary to monitor the quality of tertiary education through outcomes assessment. The change was also a response to a paradigm shift in higher education. When the focus of education moves from “instruction by the teacher” to “learning by the student”, it was deemed necessary to understand student learning through outcomes-based assessment. The national survey conducted in Japan however revealed that the assessment used might not have helped improve education (Kushimoto, 2009).

Like most of its counterparts in Asia, *Malaysia* has a very examination-oriented education system. There are four public examinations in the system – the elementary school achievement test (end of Primary 6), the lower secondary examination (end of Form 3), the Malaysian certificate of education (end of Form 5) and the higher education certificate (Form 6). Examination results are determinants of students’ progression to higher levels of education or occupational opportunities. Malaysia does have school-based assessment that aims at monitoring students’ learning growth. However, pressure on teachers to produce high test performance results in much teaching to the test and designing tests mimicking the centralized examinations. To address the growing societal dissatisfaction over the examination system,

the Minister of Education instituted several changes to improve the assessment system including placing assessment for learning as one major focus of change. In 2007, the Malaysian government recommended expanding school-based assessment and alternative assessment to provide more holistic and accurate judgments of student performance. Several challenges were perceived for successful implementation of the reform including resistance to change, the knowledge and skills of the teachers who are the assessors and the resource implications of the change (Ong, 2010).

Education in *Thailand* is centralized with a national curriculum to stipulate educational standards. Traditional paper-and-pencil tests, usually multiple-choice given at the end of learning, are normal assessment practice. The recent 1990 national curriculum states that teaching and learning activities at any level of education must emphasize “learning to think, to do and to solve problems” and that teachers must deliver instruction so as to encourage the integration of learning to know and learning to or to act (Pitiyanuwat, 2007). The Department of Curriculum and Instruction Development (CID) of the Ministry of Education is responsible for conducting a national assessment of learning outcomes at the end of elementary education (grade 6), lower secondary education (grade 9) and upper secondary education (grade 12). The aim of the assessment is to provide information for determining the standard of learning outcomes. In the classroom, teachers are advised to use formative assessment to decide the next steps for teaching, diagnostic assessment to determine what students need to improve on and summative assessment to inform the level of attainment of the students. To understand how teachers integrated assessment into teaching and learning activities, the CID conducted a pilot study in 1994. A number of assessment strategies were used including tests focusing on the skills and concepts of the subject matters and related skills, observation of practical work by the teacher, student written work, student self-assessment, and student report and records. It was found that students worked quite well in this new mode of learning. They became more self-directed. However, the CID noticed that there were some practical issues that needed attention, including providing professional training for teachers in their new roles in assessing as part of teaching, enhancing the collaboration between parents and the schools and taking actions to address large class size and teachers’ workload. For the first issue, specifically, the CID advised that teachers should be helped to develop better instructional plans and to give quality advice to students. Teachers also needed training in developing sound authentic performance tests (open-ended paper-and-pencil tests and practical tests) and marking criteria (rubrics) and in recognizing the potential for embedded assessments as part of instruction (Pravalpruk, 1999).

In *Indonesia*, the education system underwent a radical change in the twenty-first century. This reform was marked by the implementation of school-based management, which included redefining the national education objectives, decentralizing management from the government of schools and implementing the 2004 Curriculum. In the past, the Indonesian education system placed a heavy emphasis on cognitive attainment by students (Muhaimin & Ali, 2001). The new curriculum aims at promoting students’ ability in applying knowledge in real life situations and calls for teachers’ to use classroom-based assessment to support learning.

A widespread feeling is that continuous professional growth of teachers and strong school management leadership are the keys to the successful implementation of the reforms (Raihani, 2007).

In Africa, *Ghana* on the western coast of Africa had their most recent education reforms beginning in 1987 with an aim to address problems including low participation, curriculum dysfunctionality, gender disparity, rural-urban dichotomy etc. (Kwawukume, 2006). The Programme for Free Compulsory Universal Basic Education was passed by parliament in 1995 and now forms the basis of educational planning in the country. Continuous Assessment was introduced, which made the role of assessment become potentially more formative (Pryor & Akwesi, 1998). Akyeampong, Pryor, and Ghartey (2006) conducted a study investigating Ghanaian teachers' understanding of learning, teaching and assessment. It was found that the assessment teachers used was largely summative and suspected that this might result from teachers' lack of confidence and knowledge in using assessment for learning purposes. *Egypt* discussed the curriculum reform in 1993 aiming at moving children away from rote memorization and passive learning through teacher transmission, towards the model of active individual learning. To be in line with the visions of the new curriculum, assessment had to be changed (Ministry of Education, Egypt, 1995). However, the accountability and the unchallengeable rationality of the examination system left most people unable to act freely (Hargreaves, 2001). In *South Africa*, the government used continuous assessment as a means to reduce pressure from teachers and pupils but the opposite was found to be true in many schools. There was evidence to show that teacher produced tests modeling the matriculation examinations to prepare students for this high stakes university entrance examination. This increased the intensity of pressure (Lubisi & Murphy, 2002).

Generally speaking, the countries of Confucian heritage share a deep-rooted examination culture. Mainland China, Taiwan, Hong Kong, South Korea, Japan, Malaysia, Singapore (see Chapter 6 by Tan), Vietnam, Philippines and a number of other Asian countries all have examination systems that serve accountability and selection purposes. As the stakes are extremely high, schools, teachers and parents alike view preparing students for the public examinations as the ultimate goal for education. Recently, many of these countries saw the need to change this to an assessment culture that is aimed at enhancing students' all-round skills, promoting whole-person development and recognizing and developing different talents in students. Owing to their individual social, economic and educational circumstances, the countries in Asia and Africa planned and implemented their assessment reforms in their own distinctive ways but generally found tensions between the assessment reform policies and assessment practices.

7.4 Conclusion and Implications

Over centuries, assessment has been mainly used for selection and accountability purposes in the eastern and western worlds. The social and economical demands in the nineteenth century created an increasing need for trained workers of different

trades for which a standardized examination system was identified as being useful for screening and streaming purposes. In time, people became aware of the problems of high-stakes examinations and realized that, other than for selection and accountability, assessment can be used as a tool to support learning and enhance teaching. Most countries embarked on an education reform with a highly emphasised *Assessment for Learning* agenda. The highlights of this agenda include reducing excessive use of tests and examinations, and using assessment to understand and support learning, as well as using student information to improve teaching. Assessment must be consistent with the objectives of what is taught and learnt. Teachers are encouraged to use a variety of assessment strategies and assessment tasks to allow a range of different learning outcomes to be assessed. In the last few decades, there was a shift in perceptions about learning that are commonly and internationally labelled as the need for “lifelong learning”, “learning-to-learn” and “whole-person development”. Many countries highlighted in their assessment policies the need to promote learner autonomy, a key element of the above mentioned concepts. In their official documents, these governments specified the use of self- and peer-assessment to increase learners’ metacognitive abilities so that learners can take control and manage their own learning. As students’ diverse needs have got more recognized, teachers are advised to differentiate assessment strategies and tasks to identify learning needs and use them to cater for specific needs. Teachers should use assessment to develop students’ potential in different perspectives. The assessment methods and tasks to be used are varied, allowing different perspectives of learning to be facilitated and acknowledged. Basically, teachers are advised to use the information obtained to adapt teaching to the needs of the students and to change the traditional form of assessment to a more child-centred and formative one.

After years of implementation, there was evidence to show that there had been limited changes in classroom assessment practices. In general, there was over-emphasis on the grading function and under-emphasis on the learning function. The international comparison results did little to help establish an assessment for learning culture. In a number of countries, faith in assessment for learning was considerably undermined by unfavourable international comparisons. Some countries held schools and teachers accountable for the performance of their students in the standardized inter-school comparative tests. Consequently, although many teachers acknowledged the significance of formative assessment in student learning, teaching was still very much test-oriented. To help students achieve good results, a common practice was designing tests simulating the high-stakes external examination and on teaching conventional types of knowledge and competence. The above mentioned depicts a rather gloomy picture for advocates of assessment for learning reforms, as the good intentions appear to have been threatened by the worldwide dominance of high stakes summative discourse and the issues of accountability. Assessment for learning may become a major casualty of a heavily centralized education system torn between tradition and change.

The brighter side of the assessment reform movements is that the assessment landscape worldwide is gradually changing and the learning function of assessment is gaining better recognition in many education contexts. Some countries reported

success in their assessment reforms. Common to these countries are the values they see in their teachers and emphasized offering life-long professional training for teachers. Many teachers are in fact very enthusiastic about the ideas of using assessment for learning purposes. They are very willing to try out the assessment for learning concepts although generally find it rather hard to fight the examination culture and the pressure of accountability (Berry, 2010). The current problem is the widespread perception of high-stakes public examinations, believing that they are the best vehicle to boost national performances. Reviews (Black & Wiliam, 1998; Crooks, 1988; Natriello, 1987) provide clear evidence that improving the quality of formative assessment was the key to increasing student achievement. Black and Wiliam (1998) found that improvements in the quality of formative assessment resulted in effect sizes of the order of 0.4–0.7 standard deviations (equivalent to doubling the rate of learning). A more recent review of the literature on the effects of feedback and formative assessment in post-secondary education (Nyquist, 2003) found effects of similar magnitude, and, perhaps more significantly, showed that the larger effect sizes were associated with stronger implementations of the principles of assessment for learning. To improve student achievement across the curriculum, it is suggested that improving teacher quality and their capacity to use assessment as central to learning may be the most effective way to attain this goal. To make assessment a useful tool for teaching and learning, it is necessary to empower the teachers with knowledge and skills (Berry, 2011). What the teachers urgently need are, in addition to the overarching assessment policies, guidelines and directives, concrete ideas on how to translate the assessment for learning concepts into classroom actions, including, for example, detailed techniques for implementing assessment for learning in classroom situations (see Chapter 4 by Berry and Chapter 8 by Gardner et al.).

References

- Aho, E., Pitkänen, K., & Sahlberg, P. (2006). Policy development and reform principles of basic and secondary education in Finland since 1968. World Bank. http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079967208/Education_in_Finland_May06.pdf. Accessed 28 July 2010.
- Akyeampong, K., Pryor, J., & Ghartey, A. J. (2006). A vision of successful schooling: Ghanaian teachers' understandings of learning, teaching and assessment. *Comparative Education*, 42(2), 155–176.
- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University Press.
- Berry, R. (2010). Teachers' orientations towards selecting assessment strategies. *New Horizons in Education*, 58(1), 96–107.
- Berry, R. (2011). Assessment Trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice*, 18(2).
- Black, P. J. (1998). *Testing: Friend or Foe? Theory and practice of assessment and testing*. London: The Falmer Press.
- Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal*, 16(2), 249–261.
- Black, P. J., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Bonnet, G. (1997). Country profile from France. *Assessment in Education: Principles, Policy & Practice*, 4(2), 295–306.

- Brauns, H., & Steinmann, S. (1999). Educational reform in France, West Germany and the United Kingdom: Updating the CASMIN educational classification. *ZUMA-Nachrichtung*, 44, 7–44.
- Broadfoot, P. (1985). *Recent Developments in Assessment and Examination Procedures in France*. Paper presented at the Annual Meeting of the American Educational Research Association.
- Broadfoot, P. (2009). Foreword. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice*. London, New York: Springer.
- Carrasco, M. R., & Torrecilla, F. J. M. (2009). Learning assessment in Latin America. School performance behaviour and trends of Latin American pupils in primary and secondary education. *Educational Sciences Journal*, 9, 31–46.
- Council of Ministers of Education. (2005). OECD Study on enhancing learning through formative assessment and the expansion of teacher repertoires: Canadian Report. http://www.cmec.ca/Publications/Lists/Publications/Attachments/78/OECD_Formative.en.pdf. Accessed 25 July 2010.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
- Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy & Practice*, 9(2), 237–253.
- Dunleavy, J. (2007). Public Education in Canada: Facts, Trends and Attitudes. Canadian Education Association.
- Egelund, N. (2005). Educational assessment in Danish schools. *Assessment in Education: Principles, Policy & Practice*, 12(2), 203–212.
- Fernandes, D. (2009a). Educational assessment in Portugal. *Assessment in Education: Principles, Policy & Practice*, 16(2), 227–247.
- Fernandes, D. (2009b). Learning assessment in Portugal: Research and activity theory. *Sísifo. Educational Science Journal*, 9, 87–98.
- Frassinelli, L. (2006). Educational reform in Finland. <https://www.msu.edu/user/frassine/EAD845%20-%20Educational%20Reform%20in%20Finland.pdf>. Accessed 28 July 2010.
- Friesen, S. (2009). *What did you do in school today? Teaching effectiveness: A framework and rubric*. Toronto, ON: Canadian Education Association.
- Guimarães de Castro, M. H. (2001). Education assessment and information systems in Brazil. *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*. <http://www.inep.gov.br/download/internacional/idioma/Education%20Assessment%20and%20information%20systems%20in%20brazil.pdf>. Accessed 28 April 2010.
- Han, M., & Yang, X. (2001). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice*, 8(1), 5–10.
- Hargreaves, E. (2001). Assessment in Egypt. *Assessment in Education: Principles, Policy & Practice*, 8(2), 247–260.
- Hess, F. M., & Petrilli, M. J. (2006). *No child left behind*. New York: Peter Lane.
- Ingenkamp, K. (1971). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Kaser, L., & Halber, J. (2008). From sorting to learning: Developing deep learning in Canadian schools. *Education Canada*, 48(5), 56–59.
- Kiamura, K. (1997). Policy issue in Japanese higher education. *Higher Education*, 34, 141–150.
- Kushimoto, T. (2009). Outcomes assessment and its role in self-reviews of undergraduate education: In the context of Japanese higher education reforms since the 1990s. *Higher Education*, 59, 589–598.
- Kwawukume, V. (2006). *Assessment for improving learning*. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment (IAEA), Singapore.
- Lubisi, R. C., & Murphy, R. J. L. (2002). Assessment in South African schools. *Assessment in Education: Principles, Policy & Practice*, 9(2), 255–268.
- Mavrommatis, Y. (1996). Classroom assessment in Greek primary schools. *Curriculum Journal*, 7(2), 259–269.
- Ministry of Education, Egypt. (1995). *Mubarak's national project*. Cairo: MoE.

- Ministry of Education: Law 1566. (1985) Nomos Plaisio yia tin paidia no 1566. (Education Act, 1566). Athens: OEDB.
- Muhaimin, S. A., & Ali, N. (2001). *Paradigma Pendidikan Islam: Upaya Mengefektifkan Pendidikan Agama Islam di Sekolah*. Bandung: Rosda Karya.
- Na, J. (2005). The Asian Craze for Education. <http://gunsandbutter.blogspot.com/2005/05/asian-craze-for-education.html>. Accessed 28 July 2010.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155–175.
- Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis*. Unpublished master's thesis, Vanderbilt University, Nashville, TN.
- Nyström, P. (2004). *Rätt mätt på prov: om validering av bedömningar i skolan* [Validation of educational assessments] (Umeå, Umeå University, Department of Education).
- Ong, S. L. (2010). Assessment profile of Malaysia: High-stakes external examinations dominate. *Assessment in Education: Principles, Policy & Practice*, 17(1), 91–103.
- Philips, D. (2000). Curriculum and assessment policy in New Zealand: Ten years of reforms. *Educational Review*, 52(2), 143–153.
- Pitiyanuwat, S. (2007). School assessments in Thailand: Roles and achievement of ONESQA. *Educational Research for Policy and Practice*, 6(3), 261–279.
- Pravalpruk, S. W. (1999). Learning and assessment in the science classroom in Thailand. *Assessment in Education: Principles, Policy & Practice*, 6(1), 75–82.
- Pryor, J., & Akwesi, C. (1998). Assessment in Ghana and England: Putting reform to the test of practice. *A Journal of Comparative and International Education*, 28(3), 263–275.
- Raihani. (2007). Education reforms in Indonesia in the twenty-first century. *International Education Journal*, 8(1), 172–183.
- Raveaud, M. (2004). Assessment in French and English infant schools: Assessing the work, the child or the culture? *Assessment in Education*, 11(2), 193–212.
- Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *Curriculum Journal*, 18(1), 27–38.
- Sahlberg, P. (2009). A short history of educational reform in Finland. <http://192.192.169.112/filedownload/%E8%8A%AC%E8%98%AD%E6%95%99%E8%82%B2/A%20short%20history%20of%20educational%20reform%20in%20Finland%20FINAL.pdf>. Accessed 28 July 2010.
- Shimizu, Y., & Lambdin, D. V. (1997). Assessing students' performance on an extended problem-solving task: A story from a Japanese classroom. *Mathematics Teacher*, 90, 658–664.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22–27.
- Valtin, R. (2002). *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand*. Weinheim: Juventa.
- Wagner, C., & Valtin, R. (2003). Noten oder Verbalbeurteilungen? Die Wirkung unterschiedlicher Bewertungsformen auf die schulische Entwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 35, pp. 27–36.
- Wang, D. (2008). On the international influence of the Chinese imperial systems. *Hubei Normal University Journal (Educational Science Issue)*, 10(1), 71–74 (in Chinese).
- White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course. *Asian EFL Journal*, 33, 1–36.
- Wikström, C. (2006). Education and assessment in Sweden. *Assessment in Education: Principles, Policy & Practice*, 13(1), 113–128.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3/4), 283–289.