

KLE08941

**The Changing Demands of Assessment Policy:
Sustaining Confidence in Teacher Assessment**

**Val Klenowski
Queensland University of Technology**

Introduction

With the global shift towards standards-driven reform, tied to reporting, assessment issues related to public and teacher accountability take prominence. In an accountability context standards are used as a lever to improve the reliability and consistency of teacher judgement and classroom evidence is used by education systems for reporting and tracking achievement over time. Assessment is inseparable from curriculum and has become a powerful driver for change. It is central to good education and is at the heart of the teaching-learning dynamic. The relationship between the learner, learning and assessment, however, needs to be kept central and teacher assessment at the local, professional level remains fundamental.

This paper will outline the changing demands of assessment policy and associated practices for achieving accountability in the global context of standards-based reform in different education systems. What is apparent in this analysis is the central role of teacher empowerment at the local, cultural level in the case for 'intelligent accountability' (O'Neill, 2002) and more generative and educative forms of assessment, pedagogy and curriculum to enhance quality and to improve equity of educational provision. This paper argues for a central place for teacher assessment in the changing climate of assessment policy. Teacher assessment is defined as:

The process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgement, and to report at a particular time on their students' achievements (Harlen, 2005: 247).

Teacher professionalism, through educative forms of school-based and teacher-led evaluation and assessment, remains vital.

The Changing Climate of Assessment

High-stakes assessments are enacted by policy makers to improve education and, setting high standards of achievement, can inspire greater effort on the part of students, teachers and principals. Yet the inadequacy of high-stakes assessments, or the lack of sufficient reliability or validity for their intended purposes, has the potential for unintended and harmful consequences. Policy makers can be misled by 'spurious' increases in assessment results that do not relate to improved learning; students may be placed at increased risk of failure or disengagement from schooling; teachers may be blamed or

punished for inequitable resources which remains beyond their control; and curriculum and teaching can become distorted if high grades or results per se, rather than learning, become the overriding goal.

The last decade of the twentieth century saw increased international dissatisfaction with the more quantitative, traditional forms of assessment. Much of this aversion stemmed from the view of learning on which these assessments were designed and their impacts on teaching and learning. Assessment approaches from this quantitative tradition have been challenged and alternative approaches have emerged. The testing regime in England has been researched and analysed with the identification that on average each student in that education system will take over 100 formal tests or examinations (Mortimore, 2008). In October 2008 the government in England abolished national tests for 14 year olds. As expressed by the Chair of the Office of the Qualifications and Examinations Regulator (Ofqual) the government's decision to listen to advice about the impact on schools and the need for change was welcomed. She indicated how new assessments would be subject to validation by Ofqual to ensure that they command confidence, and standards are secure (International Association for Educational Assessment, 2008).

These measures parallel the changing emphases in assessment reform that include a move away from assessing knowledge and products to assessing skills, understandings and processes. Also rather than assessment occurring at the end of a course through external means assessment has been taking place throughout the course. A greater variety of methods and evidence has been sought to demonstrate learning instead of relying only on written methods and this has been accompanied by a shift from norm referencing to criterion referencing with less reliance on pass or fail, summative assessments and more attention on identifying strengths and weaknesses formatively and recording positive achievement (Torrance, 1997: 329).

Teacher Assessment

With such shifts in assessment practice the teacher assumes an important role and requires an understanding of the fundamental issues in assessment design that include 'fit for purpose' and the need for the mode of assessment to impact positively on teaching and learning. Teacher designed assessment tasks therefore need to involve a variety of contexts, range of modes within the assessment, range of response formats and styles. To achieve equity there is also a need to expand the range of indicators used to provide an opportunity for those who might be disadvantaged by one form of assessment to offer alternative evidence of their expertise. To achieve this form of assessment practice requires teacher assessment and communities of judgement practice.

Teacher assessment therefore offers an important alternative because in this context locally developed indicators can prove to be more effective educationally than examinations or tests administered from the centre. The teacher is able to attend to the student's needs that emerge from a particular context, sociocultural or historical background. One testing method does not fit all circumstances. Multiple judges are recommended and Queensland's Senior Secondary System is one such example (see the

Queensland Studies Authority website (<http://www.qsa.qld.edu.au/assessment/3111.html>). Students' work is assessed at the local level and forms part of the state system of assessment of student performance. Assessment data is collected both formally and informally and used by teachers and administrators to set learning goals and priorities to build on what students already know.

Standard-setting and assessment are linked as teachers design assessments that are intellectually challenging for their students. Teachers set standards as they identify the tasks that they want students to complete for assessment and they provide various opportunities for students to display thoughtful control over ideas.

Authentic Assessment

Alternative assessment methods emerged in response to the dissatisfaction with quantitative systems. A catalyst for such change was the realization that the type of assessment impacts profoundly on the learning dispositions, attitudes, strategies adopted and learning ability. Developments in both learning theories and the theory of educational assessment (Gipps, 1994) have supported the move towards authentic assessment.

Critique of the utility of tests in measuring what students actually know inspired a move towards 'alternative, authentic assessment approaches' (Wiggins, 1989, 1991; Newmann, 1991). Authentic assessment includes tasks that challenge the student's intellect and test intellectual ability in a manner that reflects probable experience for the individual in the field. Authentic assessment is connected to the curriculum and engages students, teachers and others in assessing performance. This form of assessment goes beyond the school for models and sites of action and promotes complex thinking and problem solving, encouraging student 'performance' of their learning and engages with issues of equity.

Authentic assessments are varied and comprehensive encouraging multiple methods for demonstrating learning. Problem solving in this assessment context requires students to think analytically and demonstrate their proficiency as they would in situations beyond the classroom. Such assessments encourage students to develop skills, understandings and insights relevant to their particular needs and contexts (for examples see the Department of Education, Training and the Arts of Queensland, Australia website <http://education.qld.gov.au/corporate/newbasics/html/richtasks/richtasks.html>).

These approaches attend to equity issues by making assessment fairer by reducing the dependence on performance in a single terminal examination as the only determinant of student achievement and by giving individuals the opportunity to demonstrate attainment over time and in a variety of contexts. This type of assessment is more accurate, and reflective of an individual's learning and development, by identifying the skills and abilities being examined. This helps to encompass a wider range of abilities and facilitates the recording of achievement.

The Queensland Curriculum, Assessment and Reporting Framework (<http://www.qsa.qld.edu.au/assessment/qcar.html>) has been developed to provide teachers with Essential Learnings (ELs), standards, Queensland Comparable Assessment Tasks (QCATs), (<http://www.qsa.qld.edu.au/assessment/4703.html>) an online assessment bank (<http://www.qsa.qld.edu.au/assessment/3162.html>) and guidelines for reporting. The purpose of the QCATs is to model quality assessment that supports teaching and learning and builds consistency of teacher judgements. These are not tests rather they aim to be authentic tasks that target a selection of ELs and incorporate standards. The emphasis is on skills such as critical thinking and reasoning involving solving meaningful problems. The QCATs are intended to provide an opportunity for students to perform their best work. Teacher assessment is fundamental to the design of the QCATs that aim to support teachers in making consistent judgements about the quality of student work and to provide information to teachers and students about what students know and can do, what is working well and what needs attention (Molloy, 2008).

Changing Demands of Assessment Policy

International Assessment Surveys

Politicians, policy makers and others are increasingly interested in international assessment surveys of educational attainment such as the Programme for International Student Assessment, (PISA) developed by the Organisation for Economic Co-operation and Development (OECD) or the Trends in Mathematics and Science Study (TIMSS) of the International Association for the Evaluation of Educational Achievement (IEA). These international comparative surveys have influenced policy development yet important questions related to design, technical issues, cultural specificity and comparability remain. The rigour of these studies, their methodologies and the way in which the data has been interpreted and used have been debated and critiqued by researchers such as (Hopmann, 2008; Wiliam, 2008; Nardi, 2008; Egelund, 2008; Wagemaker, 2008). International comparisons require a common set of criteria for measuring performance, comparability between samples and the reporting of the results, a match in terms of the content of the curriculum and/or the approach used, and regard given to context. There is a need for caution in the use of the results from such comparative studies as politicians keen to introduce particular policies on the basis of these results may overlook their limitations identified.

In fact governments have made reference to the results from international comparative surveys to justify efforts to introduce significant educational reform and curriculum change. For example, in January 2008 a National Curriculum Board was established in Australia with the Chairman of the Board, a former Director for Education in the OECD, who had been a key figure in the development of the OECD's Programme for International Student Assessment. The Federal Government's intention is that a National Curriculum in Mathematics, Science, History and English will be implemented in primary and secondary schools in 2011 and will be extended to involve geography and languages other than English in a second phase. Not surprisingly when taking up the position the Chairman of the National Curriculum Board declared in reference to his work with the OECD:

Since we began the program for the assessment of 15 year olds, countries are paying a great deal of attention. And with the release of the third set of data, last year, we can see that while Australia's maths levels are holding at the same level, we've slipped relatively because other countries have moved ahead of us. Other countries that were equal with us have moved ahead. So it's not a stable competition out there. Every country is seeking to do better and to learn from those that are doing better than themselves already. (Australian Labor Party, 2008)

Australia, as with other countries, has made use of international comparative data to introduce educational reforms or to undertake supplementary studies. The Third International Mathematics and Science Study (TIMSS) for example showed significant State and Territory differences in Australia. The analysis of international performance data has also revealed that Indigenous children scored significantly lower than non-Indigenous children (Lokan, Ford and Greenwood, 1997). In equity terms Australian schools appear not to be adequately addressing inequalities and when compared with other developed countries, Australia is underperforming in terms of equity: 'high in quality but low in equity' (McGaw, 2004).

A further example of how the results have been used in Australia to focus policy towards particular priorities is evident in an analysis of the 2003 PISA data which suggested that in general there is over-representation in the lowest categories of maths proficiency and under-representation in the highest (Thomson, Cresswell and De Bortoli, 2004: xiii). So, while the achievement of students overall is high, there are wide differences between the high and low achieving students.

Curriculum, assessment and pedagogic practices need to address such issues from a sociocultural perspective at the local professional level of the classroom. It is the professionalism of the teacher and trust in that professionalism by the political centre that will make the difference in terms of educational change for improved learning. In implementing the reforms on the basis of analyses of such comparative studies there is the danger that the heightened focus on accountability and testing will result in unintended outcomes and distrust in teacher judgement. Current methods of "accountability can damage rather than repair trust" and what is needed is *intelligent accountability* (O'Neill, 2002:58). "Trust ... is hard earned and easily dissipated. It is valuable 'social capital' and not to be squandered" (ibid: 6-7).

Impact on Policy Development

In an accountability context that increasingly demands evidence to demonstrate performance, and to illustrate the attainment of targets, an approach explicating 'what works' is pursued (CERI, 2005). As evident in the words of the Chairman of the National Curriculum Board in Australia "Every country is seeking to do better and to learn from those that are doing better than themselves already" (Australian Labor Party, 2008). Decisions regarding program and policy development have been made on the

basis of these comparisons, however, there are problems given that educational systems have different aims, and curricula, highlighting the importance for localised and contextualized interpretations of results.

Some organizations and government agencies report evidence from curriculum evaluations in terms of 'what works'. This knowledge has been described as 'a commodity that is bought, sold and applied' (Schwandt, 2003: 362). What is lost in a 'what works' approach is the value of dialogue and deliberation with practitioners in evaluation to facilitate our understanding of the challenges of diverse values in the context of practice. Rich learning comes from such discussion and an appreciation and understanding of the importance of the local context in terms of the values, aims and curricula help to identify what is valued in the particular program or educational system.

The combination of a 'what works' approach and evidence-based decision making has reinvigorated concerns relating to measurement, validity and reliability of quantitative measurement (Shaw, 1999). Government agencies are demanding Evidence Based Policy and Practice (EBPP) in this drive for efficient use of resources with the development of guidelines and frameworks to regulate and assess evaluation practice, caution must again be taken to ensure that we are comparing like with like. It is important that in the synthesis of evaluations that outcomes have not been simplified and contexts have not been ignored.

In spite of the generation of democratic, responsive and deliberative forms and purposes of evaluation it would appear that evaluation for accountability and control continues to impact on current practice as is evident in the re-emergence of bureaucratic forms of curriculum and assessment with the return of quantitative, reductionist approaches as evident in 'No Child Left Behind'. What also becomes apparent in the name of efficiency is a return to technological and behavioristic refinements of curriculum evaluation and a possible trivialization that threatens the richness of the intellectual activity for those involved in the discipline of curriculum evaluation.

The impact of such trends in evaluation on assessment practice is that data analyzed for a particular purpose may be used for another unintended purpose. For example, performance assessment data has been used for the development of league tables which are then used to judge the quality of schools. As Broadfoot (2007: 59) suggests:

...in transitions between criterion- and norm- referenced approaches and formative and summative purposes, there is considerable scope for the issue of 'fitness for purpose' to be obscured. The result ... is a number of, at best unhelpful, and at worst, downright damaging assessment practices.

Politicians have used league table comparisons to justify the introduction of additional testing regimes or performance targets without analyzing the national characteristics and contextual factors that might help teachers, parents and community members understand the difference in system-level performance.

Intelligent Accountability

In the 1980s the discourse of markets emerged in education and the place and purpose of accountability was made explicit. The dangers of 'raw' exam or test results for accountability purposes were identified. Inspection and standardized testing dominated accountability as evident in England and the main criteria for judging school performance and measuring success was student achievement. Schools are accountable for what they do for students but marketising schooling based on league tables can lead to schools being rewarded for the 'quality' of the students that they attract and enrol rather than what they actually do for students to help them achieve.

The concept of intelligent accountability includes 'good governance' that provides a level of professional freedom with some self-governance aligned to particular tasks within a framework of financial and other reporting (O'Neill, 2002). This approach allows for much of what needs to be accounted that is not always easily measured or cannot be reduced to performance indicators. 'Intelligent accountability' policies therefore involve a measure of trust-based professionalism. The five yearly inspections of schools in Finland were replaced with a requirement for schools to self-evaluate and to plan for improvement in self-identified areas. While the National Board of Education performs evaluations of Mathematics, the students' mother tongue, and may intermittently evaluate other subjects, the results are returned to schools and never made publicly available (Harris, 2006).

The Finnish system, described as an intelligent accountability system (Sahlberg, 2007), contrasts with more common systems explained as 'consequential accountability' where schools' efforts to raise performance and student achievement involve processes of promotion, inspection and the use of measures such as standardized testing to evaluate success and to determine the rewards or punishments for schools or teachers. 'Intelligent accountability' develops from a context that values and respects teachers' and principals' professionalism particularly in judging what is best for students and in reporting their achievements. Finland to date has not embraced market-oriented reforms or market-oriented management models but has focused on sustainable leadership and "built upon values grounded in equity, equitable distribution of resources rather than competition, intensive early interventions for prevention and building gradual trust among education practitioners, especially teachers" (Sahlberg, 2007: 152-53). In the Finnish education context 'intelligent accountability' enhances trust among teachers, students and education authorities in the accountability processes. What is more, they are involved in the process so they develop a strong sense of professional responsibility and initiative (Fullan, 2005). The impact on teaching and student learning has been positive.

In Finland, assessment of student learning is based on teacher assessment rather than standardized external tests, numerical grades are not used after grade five so that students are not compared with another. Grades are prohibited by law and only descriptive assessments and feedback are used which current research informs us will impact

positively on student performance and engagement in their learning (Assessment Reform Group, 1999). Teacher assessment is a dominant practice and is used as an opportunity for learning as much as for assessing student achievement.

There are shortcomings of such a system in that there is a reliance on teachers' and schools' abilities to judge and report on students' achievement and there are differences among criteria that teachers use to evaluate their students, even within the same school. Issues arise when students move to a new school and experience assessment that may involve expectations that vary to those of their previous school. Despite these shortcomings the concept of 'intelligent accountability' is preferred as it enables schools to keep the focus on learning and allows more freedom in curriculum planning compared with external standardized testing contexts. This approach allows teachers to address the needs of students from particular sociocultural contexts and enables assessment practice to be responsive.

Standards

The term 'standards' is ubiquitous and in the context of educational attainment it needs to be emphasised that there are no simple measuring instruments that can be used to determine an appropriate value for a student's achievement or for that matter of a school. There is no natural unit of measurement as there is for some physical quantities, such as weight or height. Standards are used in educational assessment and it is important to understand that the term can be used in a variety of ways. The definition of standards that is most appropriate in the context of standards-referenced assessment systems for accountability include: 'quality benchmarks' (expected practice or performance), 'arbiters of quality' (relative success or merit) and 'standards as milestones' (progressive or developmental targets) (Maxwell, 2002a:1).

Standards as 'quality benchmarks' define "an expected or typical outcome" and require representation on a continuum that defines a minimum acceptable level (Maxwell, 2008:2). Both standards as 'arbiters of quality' and 'standards as milestones' represent differentiated levels of performance. These two types of standards differ in focus and time frame so that the former may focus on a single assessment event while the latter provides for judgements over time along a continuum of learning (ibid).

The functions of standards as defined in these ways are first to provide a common frame of reference and a shared language for communicating student achievement. They are also intended to promote teachers' professional learning, focused on good assessment practices and judgement of the quality of student achievement against system level benchmarks or referents. In addition it is expected that they present more meaningful reports and engagement with assessment as a learning process.

Standards as descriptors of student achievement are used to monitor growth in student learning and provide information about the quality of student achievement. It is important to emphasise that examination or assessment standards cannot be objective in the same sense in which standards relating to physical measurements are objective.

Assessment in education is intrinsically inexact and should be treated as such (Harlen, 1994). Standards need to be described in such a way that schools can relate to them. Student work needs to be used to substantiate meaning and then the standard descriptors need to be piloted thereby grounding them in practice. They should encompass minimum and aspirational performances, be written in positive terms in language suitable for the intended audience so that moderation can occur.

Defining examination or assessment standards requires interpretation and inference fundamentally then they are subjective. The interpretation of assessment results should be in terms of being an indication of what students can do but not an exact specification (Cresswell, 2000). What should be assessed and the levels of attainment that are comparable to those represented by each grade in other examinations or assessments in the same family (Cresswell, 2000:71-72) should be defined by the standards as used in examination and assessment systems for public reporting. However, to compare attainment in different subjects we can only use indirect bases for comparison and for this we rely on statistics and expert judgement (Cresswell, 2000). Once again the role of the teacher is significant and in this context teachers have an important role in a community of judgement practice.

Teachers' Use of Standards

Most recently in Australia the National Curriculum Board has issued a proposal for discussion regarding the use of 'achievement standards'. In this context of national curriculum the term standards is used to refer to the type described as 'target benchmarks'. To illustrate:

Achievement standards indicate the quality of achievement that is expected and provide the basis for judgements about the quality of students' work. (National Curriculum Board, 2008)

The purposes that the 'achievement standards' must fulfil are outlined as follows:

- Make clear what quality of learning (knowledge, understanding and skills) is expected to be achieved;
- Provide helpful language with which teachers can discuss with students and their parents the students' current achievement level, progress to date and what should come next and
- Help identify students whose rate of progress puts them at risk of being unable to reach satisfactory achievement levels in later years (ibid).

As is apparent from current research (Klenowski and Wyatt-Smith, 2008) providing teachers with standards such as these, together with "examples of student work which provide the task and a student's response with an assessment and annotations setting out the basis for the assessment" (National Curriculum Board, 2008) will not be sufficient to achieve consistent use by teachers. For as evident both from theory (Sadler, 1987) and practice (Klenowski and Adie, forthcoming) verbal descriptors of standards that are abstract mental constructs can have their "interpretation, circumscribed, more or less

adequately, *only by usage in context*. The concrete existential referents that make up the context are essential to its proper interpretation” (Sadler, 1987: 206, my emphasis). To assess student work using standards in the form of verbal descriptors teachers will need to find the ‘best fit’ rather than the perfect match. In making such a judgement teachers will need to draw on their professional knowledge which will require contact with professional communities to create and maintain their understanding of the standards. From a sociocultural perspective it is through the interactions at moderation meetings where the justification and negotiation of the award of grades drawing on the standards (verbal descriptors), on the students’ assessed work and on the teachers’ subject and tacit knowledge that progress towards a competent assessor in a standards-referenced context will occur.

This relationship between classroom assessment and system level accountability that is focused on transparency of outcomes will be reliant on teachers’ judgements and interpretations of assessment data in the context of social moderation. For it is teachers who have direct access to the information needed for an accountability system. It is students, their teachers and their parents, who know and work with them in different settings, who are the primary sources of information to determine what schools do for students (MacBeath, 1999).

Implications for Teacher Assessment

Teachers need the freedom to make definitive evidence-based judgements on their students’ work according to established standards and a quality framework that guarantees the dependability of teacher-led assessments. The key is to use external scrutiny to maintain the quality and professionalism of teachers’ own judgements.

At the upper secondary level the assessment regime needs to reflect finer distinctions between student performance to fulfil the role of assessment for selection purposes for a wider range of destinations and progression opportunities than other levels of schooling. This is where effective and widespread use of the professional judgement of teachers is required more than ever and needs to be supported by rigorous quality assurance systems.

Moderation is one such system that serves both accountability and improvement purposes. Moderation allows for comparability of standards both within, and, between schools and an audit of range and balance in curriculum coverage is part of the process. The teacher’s role is fundamental in this process as from an analysis of the assessment data teachers develop their curriculum plans and base their teaching on the learning needs of their students.

Moderation

Moderation assists in developing coherence across the educational system. Consistency, comparability and equity are three principles relevant to moderation practice (Maxwell, 2002b). Consistency involves constancy of judgement by the individual teacher with

respect to the same evidence judged at different times and involves the equivalent application of standards across different types of evidence and opportunities for assessment. Comparability is a within-subject comparison against the performance standards for the subject. Identical aspects of knowledge, understanding and skill are not required, but equivalence of standards in terms of knowledge, understanding and skill is expected for that level of achievement. Students can be set different tasks but demonstrate a common standard of performance revealing equivalent levels of knowledge, understanding and skill.

Equity involves the opportunity for every student to reach and demonstrate their current capability. Students may demonstrate their knowledge, skills and understanding in a variety of ways so the concern should be whether they have had suitable opportunity to demonstrate what they know and can do. Moderation practice helps to ensure that these characteristics have been addressed in making judgements and that students' performances have been appropriately compared with the standard.

Moderation for accountability provides official confirmation of assessments used to report on individual students, or for cohorts of students and involves validation (Maxwell, 2002b). Validation presumes that if teachers are making appropriate judgements about a selected cross-section of student demonstrations, they will be making appropriate judgements about other student demonstrations. Moderation for accountability is designed to ensure fairness by adjusting results where there seems to be inconsistency or differences (Harlen, 1994). The moderation procedures monitor and assure comparability of the grades that are determined by this process. Important assessment data and advice are provided to teachers and schools concerning their judgements and such feedback fulfils an important quality assurance role.

Moderation for improvement involves collaborative processes promoting the professional development of teachers to undertake appropriate assessments, and to make consistent and comparable judgements (Maxwell, 2002b). It is ongoing and provides feedback for further development of comparability and may focus on both procedures and outcomes.

Research indicates that teachers who engage consistently in the moderation process are able to assess student performance more consistently, effectively, confidently and fairly. They build common knowledge about curriculum expectations and levels of achievement and can identify strengths and areas for growth based on evidence of student learning. These teachers can adjust and acquire new learning by comparing their judgements to that of others and they can share effective practices to meet the needs of all students, monitor progress, and celebrate growth (Curriculum Services Canada, 2007).

In Australia the Queensland Studies Authority (QSA) has used moderation as a quality assurance process for senior secondary studies and is currently trialing moderation practice at Years 4, 6 and 9. Moderation processes are directed at supporting and confirming understandings about judgements and performance. Teachers use assessment criteria and explicit standards to make professional judgements about performance levels demonstrated by students in the completion of assessed tasks. Teachers and assessors reach agreement

about assessments through discussion, critique and debate. They use evidence of student work to develop common understandings of the curriculum and levels of achievement to inform teaching and learning, monitoring and assessing, reporting and evaluation (Ralston & Newman, 1999).

This approach to moderation at a system level serves as vital accountability checks and balances on efforts to achieve, and demonstrate, reliability of teacher judgement in high-stakes assessment. Beyond this, however, the process of system facilitated and supported moderation provides professional development opportunities for teachers in planning teaching and learning programs, designing suitably challenging assessment tasks with accompanying statements of criteria and standards, as well as making judgements of student performance. Essentially, it is moderation that ensures that common standards are being achieved and also helps to provide comparability against benchmarks expressed as desirable features.

Increased Teacher Professionalism

Professional development occurs naturalistically through the agency of the teachers themselves as they share their knowledge and experience about working with standards in diverse school contexts and institutional settings. It is the important teacher talk and interactions during moderation meetings that impact positively on assessment practices, task design, student learning and teaching. Teacher moderation is most effective when there is “productive conflict” embedded in the school’s culture and teachers are confident to express their thinking, asking questions about the assessment data or learning after listening to others (Curriculum Services Canada, 2007). Professional learning extends beyond the time and site of the moderation meeting.

Increased professionalism, richer learning for teachers and students and more professional conversations are some of the professional benefits achievable from moderation practice. The New Basics project schools, of Queensland Australia, require students to complete ‘rich tasks’ which are carefully chosen to be intellectually challenging and to have real-world value. They are authentic. Performance on these tasks provides an informed and elaborate portrait of a student’s achievement. The evaluation of the Consensus Based Standards Validation Process of moderation used in these primary and lower secondary levels (Klenowski, 2007) found that first teacher professionalism had grown in terms of teacher confidence, building knowledge of strategies, procedures and systems to assess student work. Such teacher professional development is inherent in the process as teachers engage in rich learning conversations focused on student work and learning. The level of professional conversations increased and focused on improvement of current teaching and learning classroom practices. Teachers also gained creative ideas from a broader view of what other teachers used to achieve success. They also benefited from working and planning the assessment task together as there was richness in the learning experience and a collegial atmosphere developed with teams of teachers planning, sharing ideas and demonstrating accountability. Teachers have the most direct impact on student achievement and their role during moderation practice is fundamental.

Conclusion

Too often the policy context results in unintended consequences and unhelpful pressures on the development of assessment systems. The intended learning benefits of more productive assessment approaches are not brought to fruition; they are simply frustrated. Assessment has the potential to develop and sustain the teacher's engagement in judgement practice and curriculum planning. It is teacher assessment that has the potential to address equity issues together with the support from the political centre. In the context of changing demands of assessment policy brought about by global interest in international comparative studies and a context of heightened accountability sustaining confidence in teacher assessment remains key.

References

- Assessment Reform Group (1999) *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge, School of Education.
- Australian Labor Party (2008) *National Curriculum Board Announcement*. Press conference Queanbeyan Public School, 31 January, 2008. Retrieved November 22, 2008 <http://www.alp.org.au/media/0108/pcdpmpm310.php>
- Broadfoot, P. (2007) *An Introduction to Assessment*, London: Continuum.
- Centre for Educational Research and Innovation (CERI), (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris: OECD.
- Cresswell, M. (2000) The Role of Public Examinations in Defining and Monitoring Standards. In H. Goldstein & A. Heath, (Eds.), *Educational Standards*, Oxford: Oxford University Press. pp. 69-104.
- Curriculum Services Canada, (2007) *Teacher moderation: collaborative assessment of student work*, The Literacy and Numeracy Secretariat Capacity Building Series, Ontario: Curriculum Services Canada.
- Egelund, N. (2008) The value of international comparative studies of achievement – a Danish perspective, *Assessment in Education: Principles, Policy & Practice*, 15, 3, pp. 245-51.
- Fullan, M. (2005) *Leadership and sustainability: system thinkers in action*, Thousand Oaks: Corwin Press.
- Gipps, C. (1994) *Beyond testing: Towards a theory of educational assessment*, London: Falmer Press.

Harlen, W. (1994) *Concepts of quality in student assessment*, Paper presented at the American Educational Research Association conference, New Orleans.

Harlen, W. (2005) Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes, *Research Papers in Education*, 20, 3, pp. 245-70

Harris, J. (2006) *Alignment in Finland*, Educational Alignments Occasional Papers, Melbourne: Educational alignments

Hopmann, S. T. (2008) *No child, no school, no state left behind: schooling in the age of Accountability*, Essay, Department of Education and Human Development, University of Vienna, Austria.

International Association for Educational Assessment (2008) *UK: Government abolishes national tests for 14-year-olds*. Retrieved November 21, 2008
<http://www.iaea.info/news.aspx?id=20>

Klenowski, V. (2007) *Evaluation of the effectiveness of the consensus-based standards validation process*, Available online:
http://education.qld.gov.au/corporate/newbasics/html/lce_eval.html

Klenowski, V. and Adie, L. (in press) Moderation as Judgement Practice: Reconciling Level Accountability and Local Level Practice, *Curriculum Perspectives*.

Klenowski, V. and Wyatt-Smith, C. M. (2008) *Standards Driven Reform Years 1-10: Moderation an Optional Extra?* Paper presented at the Australian Association for Research in Education Conference, Brisbane.

Lokan, J., Ford, P. and Greenwood, L. (1997). *Math and science on the line: Australian middle primary students' performance in the Third International Mathematics and Science Study*. Melbourne, Australia: Australian Council for Educational Research.

MacBeath, J. (1999) *Schools Must Speak for Themselves*, London: Routledge.

Maxwell, G. S. (2002a) *Are core learning outcomes standards?* Brisbane, Queensland: Queensland Studies Authority. Retrieved 23 November, 2008
http://www.qsa.qld.edu.au/downloads/publications/research_qscs_assess_report_1.pdf

Maxwell, G. S. (2002b) *Moderation of teacher judgments in student assessment*, (Discussion Paper). Brisbane: Queensland School Curriculum Council.

Maxwell, G. S. (2008) *Setting standards: Fitting form to function*, Paper presented at the 34th IAEA Annual Conference, Cambridge UK.

McGaw, B. (2004). Australian mathematics learning in an international context. In I.

Putt, R. Farragher, and M. McLean (Eds.), *Mathematics Education for the Third Millennium: Towards 2010*. Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia. (p. 29). Melbourne: MERGA.

Molloy, B. (2008) *An Update on QCATs and Assessment Resources – Extended Trial*, Queensland Studies Authority, Brisbane.

Mortimore, P. (2008) *Australian Education Reforms: Lessons from England*, Presentation Brisbane.

Nardi, E. (2008) Cultural biases: a non-Anglophone perspective, *Assessment in Education: Principles, Policy & Practice*, 15, 3, pp. 259-66.

National Curriculum Board (2008) *The Shape of the National Curriculum: A Proposal for Discussion*, Retrieved 5 November, 2008
http://www.ncb.org.au/our_work/preparing_for_2009.html

Newmann, F. (1991) Linking restructuring to authentic student achievement, *Phi Delta Kappan*, 72, 6, pp. 458-63.

O'Neill, O. (2002) *A Question of Trust: The BBC Reith Lectures*, Cambridge: Cambridge University Press.

Ralston, F & Newman, H. (1999) *Towards consistency of teacher judgment: moderation in all things*. Paper presented at Australian Curriculum Studies Association Conference, 29 September – 2 October, Perth.

Sadler, D. R. (1987) Specifying and promulgating achievement standards, *Oxford Review of Education*, 13(2), 191-209.

Sahlberg, P. (2007) Education policies for raising student learning: The Finnish approach, *Journal of Education Policy*, 22, 2, pp. 147-171.

Schwandt T. A. (2003). 'Back to the rough ground!' Beyond theory to practice in evaluation. *Evaluation*, 9, 353-364.

Shaw, I. (1999). *Qualitative Evaluation*. London: Sage.

Thomson, S., Cresswell, J. & De Bortoli, L. (2004) *Facing the future: A focus on mathematical literacy among Australian 15 year-old students in PISA 2003.*, Camberwell: Australian Council for Education Research.

Torrance, H. (1997) Assessment, accountability and standards: Using assessment to control the reform of schooling. In A.H. Halsey, H. Lauder, P. Brown & A.S. Wells (eds) *Education Culture, Economy and Society*, Oxford: Oxford University Press.

Wagemaker, H. (2008) Choices and trade-offs: reply to McGaw, *Assessment in Education: Principles, Policy & Practice*, 15, 3, pp. 267-78.

Wiggins, G. (1989) A true test: Toward more authentic and equitable assessment, *Phi Delta Kappan*, 70, 9, pp. 703-13.

Wiggins, G. (1991) Standards, not standardization: evoking quality student work, *Educational Leadership*, 48, 5, pp. 18-25.

Wiliam, D. (2008) International comparisons and sensitivity to instruction, *Assessment in Education: Principles, Policy & Practice*, 15, 3, pp. 253-57.