

**The Role and Purpose of Standards in the Context of National Curriculum and
Assessment Reform for Accountability,
Improvement and Equity in Student Learning
Val Klenowski, QUT & Claire Wyatt-Smith, Griffith University**

There is an already strong and growing policy and public interest push at both national and international levels for public institutions, including corporations, hospitals, government agencies and schools, to be held accountable. In the case of education policy, this has led to heightened interest and considerable investment in implementing external accountability systems, tied in many cases to large-scale standardised testing and public reporting. As evident in several countries, including the UK and USA, the education reform moves associated with consistent curriculum (irrespective of whether at state or national levels), standardised testing and reporting for public accountability purposes has reinforced the oft-reported tension between information to improve teaching and learning, and information to inform the public of education quality. As Shavelson, Black, Wiliam, and Coffey warn,

While polls show widespread support for the noble democratic concept of accountability, accountability can and does fall short in practice. When the stakes are high, as they are now in education accountability systems, and when the interpretations of large-scale assessment scores with ambiguous or narrow meaning are treated in league tables and funding decisions as unambiguous, and when single scores are generalised beyond justification as true characterisations of individuals and systems, the potential for mischief is enormous. (2004, p. 35)

Our interest in this paper is not to fuel the tension around the longstanding distinction between assessment for improvement and assessment for measurement. Instead, it is to consider how interpretations of assessment scores, as referred to above, highlight the need to understand both the nature and purposes of standards. Of particular interest is the 'fit' between how standards are formulated and how they are used in practice, by whom and for what purposes.

Within the current context of the development of a national curriculum in Australia, while there is a focus on curriculum and achievement standards, what has been missing from the public articulation of developments, is the identification of the assessment evidence that will be used in the formation of the standards. In fact there has been limited information about how the national achievement standards will be developed, by whom and how they will be used in practice.

Our starting proposition is that, if the unintended consequences of previous curriculum and assessment reforms are to be avoided then being clear about the role of standards in national curriculum and assessment reform for the purposes of accountability, improvement and equity is essential. This paper investigates the evidence relating to standards and the relationships to curriculum, accountability, improvement and equity. In particular the paper addresses the following questions:

From the research related to the functions and purposes of ‘standards’ what evidence is there that supports teachers’ use of standards for system reporting as valid and reliable practice?

What are the conditions required for standards to not only be used to measure improvement but also to inform all student learning and teaching of all students for improvement purposes in the context of national curriculum and assessment reform?

Background

In 2007 the six states and two territories of Australia developed individual approaches to the use of standards in the implementation of curriculum, assessment and reporting. However with the establishment of the National Curriculum Board (www.ncb.org.au) in February 2008 the curriculum and assessment context began to change. This Board was established to set the *core content* and *achievement standards* in Mathematics, Science, History and English from Pre-school to Year 12. There are plans now for a national curriculum to be developed by 2010 and in 2011 a second phase will extend the development to include Languages and Geography. The rationale for these developments was derived partly from an earlier investigation of the introduction of an Australian Certificate of Education aimed at achieving greater consistency in senior secondary arrangements for curriculum, assessment and certification, more comparable student results across Australia, and more consistent standards of student achievement (Masters, Forster, Matters, & Tognolini, 2006).

A further study (Matters & Masters, 2007) was subsequently commissioned to investigate what was common content, what was essential curriculum content and whether achievement standards were comparable in the final year of schooling, in English (including Literature), Mathematics, Chemistry, Physics and Australian History. This study found that there was significant consistency in what was assessed, however, different jurisdictions use different methods of assessment (e.g. external examinations, teacher-devised assessment instruments and projects). This raised the important question as to whether achievement standards can be compared across jurisdictions, or whether the existence of different assessment methods hinders comparison. The study recommended that for each nominated senior school subject a curriculum ‘core’ be identified that clearly specified what all students taking that subject would be expected to learn, regardless of where in Australia they live. It also recommended that a set of achievement standards be developed as a nationally consistent description of how well students are expected to learn the core in each subject.

While the report indicated that the call for greater consistency, increased comparability, and clearly stated achievement standards did not necessarily imply the need for a national curriculum or common national subject examinations, it noted that achievement in these areas was inevitably more difficult if the underlying certificates, curriculum and assessment programs were to be independently developed and managed (Matters & Masters, 2007).

Current Context

By May 2009 the National Curriculum Board had through a process of consultation, managed the development of four framing papers in the subject areas of English, Mathematics, Science and History. This work was handed over to the newly established Australian Curriculum, Assessment and Reporting Authority (ACARA), an independent statutory authority. ACARA now has the responsibility for the management and the implementation of the national curriculum (to be referred to henceforth as the Australian Curriculum), national student assessment and reporting of school education outcomes. There is also an intention to establish a standards-referenced framework to “invigorate a national effort to improve student learning in the selected subjects” (National Curriculum Board, 2008, p. 3).

Definitions

In pursuing the first question that this paper addresses related to the functions and purposes of ‘standards’ and the evidence that supports teachers’ use of standards for system reporting as valid and reliable practice, it is important to understand the different distinctions in the definitions of standards. The meaning of the term will vary according to the context in which it is used and also the purpose (goal) and function (role) it fulfils.

The word ‘standard’ is ubiquitous yet difficult to define for its meaning is derived from its historical and social context, consequently different countries have varying views about what constitutes a ‘standard’ (Goldstein & Heath, 2000). Dictionary definitions also illustrate that the term has different meanings and that they change with time and will continue to do so. The concept of standards is therefore elusive and confusion can often occur when the term is used in official documents or when making comparative judgments because it is not always clear which meaning is intended.

To begin, the distinction between content and performance or achievement standards needs to be stated as they are often used in the context of assessment. Content standards apply to schools and systems and generally refer to the knowledge and/or processes that are taught. Maxwell (2009) emphasises that these standards help schools develop their curriculum in relation to their local contexts. Performance or achievement standards apply to students and refer to what they have learned. They are usually embedded in the tasks that the students need to complete drawing on their knowledge and skills (Marsh, 2009). They are used for assessing summatively and to report on the quality of the achievement or performance of the student. These standards can also be used formatively to inform students of their strengths and areas for development (Maxwell, 2009). In what follows we argue that, in the context of national curriculum and assessment reform, what has been learnt is the need for four conditions to be addressed when implementing standards. These conditions are outlined next.

First Condition: Purposes and Functions of Standards

The first condition is to be clear about the **purposes** of standards and their **functions**. This is important in a context where there is a growing global trend for using standards not just for accountability but also for the purpose of improving learning.

In Australia the term standard is used in different contexts to fulfil different roles. First is an example of the term’s use in relation to one of the functions of ACARA which is to

“develop and administer a national school curriculum, including content of the curriculum and *achievement standards*, for school subjects specified in the Charter” (Australian Government, 2008, p. 5, our emphasis). Achievement standards have been defined as follows:

43 Descriptions of content make clear what should be taught and the knowledge, understandings and skills that students are to acquire or develop. Achievement standards indicate the quality of achievement that is expected and provide the basis for judgements about the quality of students’ work.

44 It is not helpful to describe a single achievement standard for a year level (or the end of a phase/stage) because of the wide variation in students’ achievements. Focusing on a minimum benchmark is unhelpful for students who might have reached and surpassed it in earlier years of schooling and is not necessarily all that helpful for students who might not have reached the benchmark but whose achievement levels are improving year on year. (National Curriculum Board, 2008, p. 9)

Most recently ACARA has stated that the achievement standards for K – 10 will be represented at every year of schooling by: a statement of learning typically expected for the year; a set of generic grade descriptors and a set of work samples illustrating the quality of expected learning. The use of annotated student work samples aims to illustrate the differences in quality of student work. The achievement standards are intended to provide “... an expectation of the quality of learning that students should typically demonstrate by a particular point in their schooling (that is, the depth of their understanding, the extent of their knowledge and the sophistication of their skills)” (ACARA, 2009, p. 20).

These purposes of the ‘achievement standards’ then include first, to make clear the expected quality of learning (knowledge, understanding and skills) to be achieved, second, to provide helpful language with which teachers can discuss with students and their parents the student’s current achievement level, progress to date and what should come next, and third, to help identify students whose rate of progress puts them at risk of being unable to reach satisfactory achievement levels in later years (National Curriculum Board, 2008). These standards are intended to fulfil the purpose of improvement of student learning and accountability.

In the following example the term is used in the context of the National Assessment Program Literacy and Numeracy (NAPLAN) testing and fulfils a particular role in this context

For each year level a *national minimum standard* is located on the scale. For Year 3 Band 2 is the national minimum standard, for Year 5 Band 4 is the national minimum standard, for Year 7 Band 5 is the national minimum standard and for Year 9 Band 6 is the national minimum standard. The skills that students are typically required to demonstrate for the minimum standard at each year level are described on the back page of the student report.

These *standards* represent increasingly challenging skills and require higher scores on the national scale. (Ministerial Council on Education, Employment, Training and Youth Affairs, 2009, our emphasis)

In some states of Australia, such as Queensland, the state government is keen to raise standards as represented by the results of NAPLAN testing. For example, in 2009 the premier advised schools to sit practice NAPLAN tests in Years 3, 5, 7 and 9 as she was disappointed by the overall results of the 2008 tests which that she indicated were designed to assess if students were meeting “national *standards* in numeracy, reading, writing, spelling, punctuation and grammar” (Bligh, 2009, our emphasis). What is confounding for teachers is that currently in Australia there are no statements about the expected learning of literacy and numeracy and no standards to inform them about the expectations of quality. There are only summary statements of skills assessed to inform parents about their child’s report.

The distinctions, in the uses of the term standards, need to be made explicit. In the first example the term standard is used in the context of ACARA as achievement standards. The notion of a standard in this context is as a measure or yardstick for judging achievement. In the second example the term is used in reference to national minimum standards and the Queensland’s premier’s response to the NAPLAN testing program highlights how the meaning of the term standard differs in that it is used as a level of attainment or point of reference as measured by a yardstick, or as in this case, band levels on a scale. The concern for teachers is that by emphasising that the NAPLAN test is the measure or reference point, the consequent action by teachers will be to narrow their focus to that which is tested or measured. In other words the curriculum too will be narrowed and teachers will emphasise in their teaching that which has been specified in the test.

As is also evident from the Queensland government’s response to the NAPLAN results, governments are increasingly anxious about education standards particularly as reflected in such national or international comparisons of student achievement, because of the expected critical contribution to economic growth and competitiveness. There is also increasing individual (particularly parental) anxieties because of the growing importance of formal qualifications in determining success in terms of life chances.

Standards for Improving Learning

Achievement standards are intended for the purpose of indicating the quality of achievement that is expected and provide the basis for judgments about the quality of students’ work. The purpose here is to use the standards in relation to the improvement of student learning.

Research indicates that standards are useful for the **purpose** of informing teachers’ work and in contributing to quality teaching and learning experiences (Klenowski, 2006, 2007; Sadler, 2005; Wyatt-Smith & Castleton, 2004). Standards, as descriptors of student achievement, **function** by monitoring the growth in student learning and by providing information about the quality of student achievement to fulfil the purpose of

improvement in student learning. Standards used to assess the quality of learning assist teachers in identifying areas for improvement in teaching, curriculum design or development. The provision of these standards that can make explicit for teachers what to teach and the level of performance expected for a particular age group and in this way can assist in meeting the demand for public accountability at the local professional level of the teacher (Harlen, 2005; Wilson, 2004).

Examples of standards for improvement of student learning are the standards for the Essential Learnings, which provide a generic description of the expected quality of student work, and provide a common language for teachers to use in discussing student work (Queensland Studies Authority, 2007), the standards used to assess the Queensland Comparable Assessment Tasks (QCATs) are a further example. These standards are also intended to promote teachers' professional learning, focused on good assessment practices and judgment of the quality of student achievement against system level benchmarks or referents. In addition it is expected that teachers using the standards will present more meaningful reports and engagement with assessment as a learning process.

Standards for Accountability

The standards as described in the NAPLAN testing example fulfil an accountability function. For the **purposes** of accountability standards defined as 'quality benchmarks' (expected practice or performance), 'arbiters of quality' (relative success or merit) and 'standards as milestones' (progressive or developmental targets) (Maxwell, 2002, p. 1) seem most appropriate. Standards as 'quality benchmarks' describe "an expected or typical outcome" and require representation on a continuum that defines a minimum acceptable level (Maxwell, 2008, p. 2). Standards, as 'arbiters of quality' and 'standards as milestones', represent differentiated levels of performance. The **roles** or the way that these representations fulfil the purpose of accountability is by the mode of representation as 'benchmarks', 'arbiters' or 'milestones'. The difference occurs in terms of focus and time frame so that standards as 'arbiters of quality' function by the focus on a single assessment event, while the standards as milestones function by providing for judgments over time along a continuum of learning. Standards as defined in these ways provide a common frame of reference and a shared language for communicating student achievement. Standards need to be described in such a way that schools can relate to them.

Governments and policy officers enact high-stakes assessments and set high standards of achievement to improve education to inspire greater effort on the part of students, teachers and principals. The inadequacy of high-stakes assessments, in their lack of sufficient reliability or validity, for their intended purposes can result in unintended consequences. To illustrate, increases in assessment results may **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 that is beyond their control; and curriculum and teaching can become distorted if high grades per se become the overriding goal (Klenowski, 2009). The No Child Left Behind Act (NCLB, 2001) in the USA is an example where the push to raise standards has led to enormous

pressure on teachers and distortions in the teaching of a holistic curriculum and the reduction in authentic and challenging learning experiences for students (Marsh, 2009).

Second Condition: Understanding the Representations of Standards

This second condition relates to the teacher's understanding of the representation of the standards. What becomes apparent is that no matter the context or the purpose of standards, professional judgments are needed to describe and maintain standards and this implies a degree of trust of the professionals. For as described by Goldstein and Heath,

It is difficult, if not impossible, to arrive at an 'objective' definition of educational standards. Despite claims to the contrary, ultimately the final appeal is to human judgement and no amount of technical sophistication can alter this. (2000, p.8)

Trust between educators and the public is therefore a recurring topic in relation to the use of standards in curriculum reform. Policies based upon comparisons of examinations, tests or other devices should therefore be seen for what they really are, human judgments. However conscientiously pursued they are ultimately subjective, or reflective of the individual's view, and are influenced by culture, personality and general perceptions of the external world (Goldstein & Heath, 2000, p. 8).

Defining examination or assessment standards requires interpretation and inference so fundamentally they too are subjective or reflect the individual teacher's perception. The interpretation of high stakes tests or examination 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) 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 judgment (Cresswell, 2000).

In the context of examinations, high stakes testing or in the use of standards for improving learning the teacher has an important role in a community of judgment practice. This is because standards-referenced assessment relies on teacher judgment that can be made dependable if standards are promulgated in appropriate forms and teachers have the conceptual tools and professional training. Teacher judgment is central to the use of standards and moderation.

Standards are understood differently dependent on the **context and their purpose**. The different representations and models of standards therefore need to be defined and understood in relation to the context and the purpose for which they are used. The methods of promulgation include: numerical cut-offs, tacit knowledge, exemplars and verbal descriptors (Sadler, 1987).

Artefacts such as exemplars or model answers can also represent standards.

Exemplars help to explicate judgment practice and form one part of a comprehensive approach to moderation. Not only are annotated samples of each standard (A-E) required but also an overall commentary for each, detailing the approach used to reach the judgment (i.e. holistic, analytic, trade-offs etc). There is a need to improve and support

judgment practice through the provision of exemplars. How the features of the standard are communicated can have an important effect on teaching and student learning.

Third Condition: Moderation Practice

Our third condition relates to opportunities for teachers to share interpretations of assessment criteria and standards through social or consensus **moderation**. We invite readers to consider how consensus moderation is key to system efforts in building teacher assessment capacity, as well as teacher confidence in the judgments they make of student work. Beyond this, efforts to use teachers' judgments of student achievement for the purposes of local level assessment and system level accountability necessarily require 'a way to integrate teachers' judgments of students' responses to the various assessment modes with those of other teachers' (Wilson, 2004, p. 11). This process is called moderation and is described as a 'form of quality assurance for delivering comparability in evidence-based judgments of student achievement' (Maxwell, 2007, p. 2).

Maxwell highlighted two functions of moderation namely quality assurance and comparability, though as we suggest below, moderation can benefit curriculum design and delivery in the classroom:

- *Quality assurance* refers to methods for establishing confidence in the quality of procedures and outcomes.
- *Comparability* requires assessment against common characteristics or criteria, such as provided by a subject syllabus or other frame of reference and requires consistency in the application of common standards so that all achievements given the same grade or level of achievement have reached the same standard. (Maxwell, 2007, p. 2)

Several writers (Harlen, 2005; Sadler, 1989; Wyatt-Smith & Castleton, 2005; Wyatt-Smith, Klenowski & Gunn, in press) have emphasised how common standards provide external reference points for informing judgment and are pivotal for achieving comparability. Consensus moderation means that the frames of reference (standards, scoring guidelines, assessment criteria etc) must be defined and disseminated to allow for common interpretation (Maxwell, 2007, 2009). This observation calls for clear recognition of the social nature of moderation where teachers interact with one another, sharing judgments of student work samples. Such sharing is an act that necessarily involves an openness to making available information about interpretations of the standards, disclosures that otherwise remain private and unarticulated.

In addressing ways to achieve high reliability while preserving validity, several writers have argued that it is important for teacher assessors to develop common understandings of mandated standards and reach 'similar recognition of performances that demonstrate those standards' (Maxwell, 2001, p. 6). However, clear communication about the nature of standards and the levels they seek to specify is not necessarily achieved through the provision of stated standards alone. Sadler (1989) argued, for example, that exemplars or samples of student work provide concrete referents for illustrating standards that otherwise remain abstract

mental constructs. He made the point that the stated standards and exemplars work together to show different ways of satisfying the requirements of say, an A or C standard. A related point is that standards written as verbal descriptors call for qualitative judgments. As such, the standards necessarily remain open to interpretation and common understandings of the terms used to capture notions of quality in the standards need to develop over time.

Given that standards require interpretation, moderation provides the means through which teachers meet to review how they have interpreted and applied given standards, and in this way moderation is vital in system efforts to promote a more consistent use of standards over time and across the country.

While statistical moderation is widely practised, consensus moderation is not well understood internationally. More specifically, there are few examples in education policy internationally of teachers being involved in moderated school-based assessment where they come together to use standards to inform their decisions about the quality of student work they are assessing. The Queensland system of externally moderated standards-based assessment in senior schooling¹ is an example of note, as is the Queensland Curriculum Assessment and Reform initiative², which has recently attempted standards-referenced moderation in Years 1–9. While the details of these two approaches necessarily differ, common to them is the understanding that conditions can be installed at system level to ensure teachers reach judgments with high validity and high reliability levels.

This is not to suggest that the function of moderation is narrowly understood as serving accountability alone. Instead, the proposition we are offering is that moderation practices where teachers come together to assess and judge student work against stated standards can have a direct flow on to and benefit for system efforts to improve curriculum design and development in the classroom. Specifically, it is in the context of standards-based moderation talk that teachers can explore the meaning and application of standards as this relates to construct validity, achieving clarity of expectations for themselves and their students in relation to task design. Further, moderation can function as a main means through which teachers reach agreement regarding the qualities of the learning being assessed.

Fourth Condition: Assessment Community

Our fourth condition relates to the need to recognise influences on judgment and how those influences are dependent on assessment purposes. For example, there are assessment purposes that involve judgment for accountability, and reporting of student achievement more specifically. These can be readily distinguished from assessment and judgements that have as their primary concern the improvement of learning. By way of illustration, a teacher may adapt the curriculum and assessments for students with learning difficulties, providing opportunities for those students to achieve goals that are

¹ For more information visit the QSA website: www.qsa.qld.edu.au/assessment/2130.html

² For more information visit the QSA website: www.qsa.qld.edu.au/assessment/qcar.html

realistically attainable for them, though below the goals of other students. Such adaptations permit learning and assessment to be tailored to the interests and needs of individual students and as such, they do not necessarily reflect standards expected of students at a given year level. They do however serve the purpose of promoting and monitoring learning for individuals who, over time, may well achieve those year level standards.

A word of caution applies therefore to the notion that standards necessarily should have a regulatory influence over teaching. Instead, as mentioned above, they can serve to inform teachers about curricular intent and the demands of assessment tasks relative to that intent.

Further, while standards and exemplars together can serve to make clear desired characteristics of quality, they do not necessarily account fully for the factors that shape teacher judgment. In a three-year large-scale Australian study of teacher judgment in middle schooling, Cooksey, Freebody, and Wyatt-Smith (2007) reported high levels of variability in teachers' notions of quality and also unearthed the range of factors that shape how judgments are reached (Wyatt-Smith et al, 2008). Similarly, **in another study** Wyatt-Smith et al. (in press), found that while teachers take account of centrally developed and mandated standards, their judgment acts, as displayed in recorded moderation sessions, go well beyond the use of standards. Specifically, these researchers identified that while teachers did use stated standards and related textual resources (e.g., sample responses and Guide to making judgments), they also actively referred to other tacit knowledges (e.g., teachers' personal knowledge of students; knowledge of curriculum and teaching contexts where they have delivered the curriculum; prior evaluative experience and tacit or in-the-head standards not elsewhere specified) for judgment purposes. Parts of this second category of resources were often used in combination, and sometimes in opposition to the stated standards. They reported how, at times, the other knowledges were used as a reason for discounting, or even subverting the stated standards. Given this, it is crucial that practical, unambiguous guidelines and professional development opportunities be provided to teachers about desired judgment practice and the legitimacy (or otherwise) of the various resources available for teachers to draw upon.

These observations point to the need for giving priority to the development of an assessment community within a school. Gunn (2007) commented on this as follows:

Perhaps the most clamorous challenge is ongoing support for teachers in the form of appropriate professional development and preservice education. Any change to assessment processes hinges on support *from* teachers, and support *for* teachers, to ensure an ability to adapt to and consider changes at the classroom level. Such changes ideally should follow deep professional knowledge and understanding of the system to be implemented. The challenge is to ensure that teachers have the requisite opportunities for working collaboratively across, as well as within, schools. Concurrent with this are opportunities for gaining deep understanding of the theoretical underpinnings of standards-referenced assessment along with knowledge of assessment practices including task design and key features of

effective moderation with a focus on standards. This requires theoretically-based yet practically-situated learning rather than decontextualised one-shot professional development. (p. 59, our emphasis)

Conclusion

This paper puts forward a proposal for reviewing the role and purpose of standards in the context of national curriculum and assessment reform more generally. It seeks to commence the much needed conversation about standards in the work of teachers (as distinct from large scale testing companies and policy personnel responsible for reporting) and identifies some key conditions that relate to effective use of standards to measure improvement and support learning.

While there has been considerable investment of energy in discussions and forums about national curriculum in Australia, there has been a striking silence about how assessment of student achievement in the national curriculum will occur. Also striking has been the limited public attention given to the standards that might apply for gauging student achievement in the new curriculum. Yet, at another level it is perhaps consistent with other curriculum policy initiatives where assessment remains unaddressed until after curricular decisions are taken. This approach necessarily continues the long and unhelpful tradition of separating curriculum and related teaching and learning activities from assessment.

The overview of the conditions presented in the paper points clearly to the need for building the capability of the workforce, if educational assessment policy is to engage the profession with realising the potential of standards to inform teacher judgment and in turn, improve student learning and outcomes. As a corollary of this, we suggest that improvement will not come from curriculum reform in and of itself, and that it is timely to review the role of teachers as the primary assessors of student learning. In this role, the country's prospects for achieving improved learning and indeed, greater equity opportunities in schooling, are directly tied to efforts to achieve improved assessment literacy on the part of policy officers, teachers, principals and educators in general. In part, this can be achieved through pre-service and in-service development with a focus on quality assessment practices including the use of standards, evidence through case studies of informed practice. It could also be achieved through a greater balance in the policy direction to promote the improvement function of standards, a focus that can be often lost in the intense policy interest in standards for reporting purposes alone.

References

- Australian Curriculum, Assessment and Reporting Authority. (2009). *Curriculum design paper*. Retrieved 16 June, 2009, from www.acara.edu.au/verve/resources/Curriculum_Design_Paper.pdf
- Australian Government. (2008). *Australian Curriculum, Assessment and Reporting Authority Act 2008, No. 136, 2008*. Retrieved 12 June, 2009, from [www.comlaw.gov.au/ComLaw/Legislation/Act1.nsf/0/8F3E822CC1A33B3DCA25751B007E3618/\\$file/1362008.pdf](http://www.comlaw.gov.au/ComLaw/Legislation/Act1.nsf/0/8F3E822CC1A33B3DCA25751B007E3618/$file/1362008.pdf)
- Bligh, A. (2008) Letter to Parent, Brisbane: Queensland Government.
- Cresswell, M. (2000). The role of public examinations in defining and monitoring standards. In H. Goldstein & A. Heath (Eds.), *Educational standards* (pp. 69-104). Oxford: Oxford University.
- Gunn, S. J. (2007). Literature review. In C. Wyatt-Smith & G. Matters, *Proposal for a new model of senior assessment: realising potentials*. Unpublished report prepared for Queensland Studies Authority, June 2007.
- 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), 245-70.
- Goldstein, H., & Heath, A. (Eds.). (2000). *Educational standards*. Oxford: Oxford University.
- Klenowski, V. (2006). *Evaluation report of the pilot of the 2005 Queensland Assessment Task (QAT)*. Brisbane: Queensland Studies Authority. Retrieved from: <http://www.qsa.qld.edu.au/research/reports.html>
- Klenowski, V. (2007). *Evaluation of the effectiveness of the consensus-based standards validation process*. Retrieved from: http://education.qld.gov.au/corporate/newbasics/html/lce_eval.html
- Klenowski, V. (2008). *The changing demands of assessment policy: Sustaining confidence in teacher assessment*. Paper presented at the Australian Association of Research in Education, Brisbane, Queensland.
- Marsh, C. (2009). *Key concepts for understanding curriculum*. London: Routledge.
- Masters, G.N., Forster, M., Matters, G.N., & Tognolini, J. (2006). *Australian Certificate of Education: A way forward*. Melbourne: Australian Council for Educational Research.
- Matters, G. N., & Masters, G. N. (2007). *Year 12 Curriculum Content and Achievement Standards*. Brisbane: Australian Council for Educational Research.
- Maxwell, G. S. (2001). *Are core learning outcomes standards?* Brisbane: Queensland Studies Authority.
- Maxwell, G. S. (2002) *Are core learning outcomes standards?* Brisbane, Queensland: Queensland Studies Authority. Retrieved 23 November, 2008, from www.qsa.qld.edu.au/downloads/publications/research_qsc assess_report_1.pdf
- Maxwell, G. (2007). *Implications for moderation of proposed changes to senior secondary school syllabuses*. Paper commissioned by the Queensland Studies Authority. Brisbane: Queensland Studies Authority.
- Maxwell, G. S. (2008). *Setting standards: Fitting form to function*. Paper presented at the 34th IAEA Annual Conference, Cambridge UK.
- Maxwell, G. S. (2009). Defining Standards for the 21st Century, in C.M. Wyatt-Smith &

- J. J. Cumming (eds.), *Educational Assessment in the 21st Century*, Springer?
- Ministerial Council on Education, Employment, Training and Youth Affairs. (2009). *National Assessment Program Literacy and Numeracy*. Retrieved 12 June, 2009, from www.naplan.edu.au/frequently_asked_questions/frequently_asked_questions.html
- No Child Left Behind Act. Reauthorization of the Elementary and Secondary Education Act. Pub. L. 107-110 2102(4) (2001).
- National Curriculum Board (2008) *The Shape of the National Curriculum: A Proposal for Discussion*. Retrieved 5 November, 2008, from www.ncb.org.au/our_work/preparing_for_2009.html
- Queensland Studies Authority. (2007). *Information statement April 2007: Standards draft 2*. Brisbane, Australia: QSA.
- Shavelson, R.J., Black, P.J., Wiliam, D., & Coffey, J. (2004). *On linking formative and summative functions in the design of large-scale assessment systems*. Retrieved 3 May, 2007, from www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/On%20Aligning%20Formative%20and%20Summative%20Functions_Submit.doc
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, 30(2), 175-194.
- Wilson, M. (Ed.) (2004). *Towards coherence between classroom assessment and accountability*. Chicago: University of Chicago Press.
- Wyatt-Smith, C., & Castleton, G. (2004) Factors affecting writing achievement: Mapping teacher beliefs. *English in Education*, 38(1), 37-61.
- Wyatt-Smith, C., & Klenowski, V. (2008). *Examining how moderation is enacted within an assessment policy reform initiative: You just have to learn how to see*. Paper presented at the 34th International Association for Educational Assessment Annual Conference, Cambridge, England, September 2008..
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. J. (in press). The centrality of teachers' judgement practice in assessment: a study of standards in moderation. *Assessment in Education*.