Chapter 14
Defining Standards for the 21st Century

Graham Samuel Maxwell

Introduction

Most countries around the world have entered the 21st century with increased focus on and requirements for educational accountability, expressed through a variety of assessment regimes and policies. Common to these directions is talk about 'standards' in education. For example, there is talk about setting standards (preferring high standards and eschewing low standards), monitoring standards (emphasising school and teacher accountability), raising standards (improving educational outcomes) and reporting on standards (saying how well students are progressing in school). Talk about standards pervades current discussions about education, particularly, for the focus of this book, discussions that involve educational assessment.

But what exactly are standards and how are they expressed? Discussions about standards are not all about the same thing. The term 'standards' has a variety of meanings in different contexts and different countries. These different meanings can have quite different implications for educational practice. Clear communication depends on identifying which meaning is intended. However, there is no agreed conceptual structure for identifying different kinds of standards, the ways in which they are expressed and their consequences or effects. In order to support meaningful educational theories and practices both within and across countries in the 21st century and to address educational issues on a global basis, there is an urgent need to dispel some of the confusion surrounding standards and to develop some clearer conceptual structures.

In this chapter, I examine different perspectives on standards and suggest some ways in which to clarify their different meanings and uses. I draw some distinctions between different kinds of standards, especially those relating to expected student learning outcomes. I discuss how standards are represented or expressed, consider some unresolved issues and suggest some desirable directions of development for the 21st century.

G.S. Maxwell (✉)
Educational Consultant, Brisbane, QLD, Australia
e-mail: gsmax@bigpond.net.au

C. Wyatt-Smith, J.J. Cumming (eds.), *Educational Assessment in the 21st Century*,
DOI 10.1007/978-1-4020-9964-9_14,

The chapter is presented in three sections, with the first discussing the dimensional meanings of standards and the second analysing some of the current discourse around standards. The final section considers some current educational realisations and uses of standards, and their import for education and students.

While the focus of this discussion is on different meanings of standards and their implications for educational practice, I also explore the consequences of different realisations of standards, especially from the perspective of implications for the student as learner. My underlying philosophical position or assumption is that the purpose of education is to enable the advancement of the personal knowledge and capabilities of each student to the fullest extent possible and to prepare them for further learning and development throughout their life. For children just beginning school now, this can mean throughout the whole of the 21st century.

Charting the Different Meanings of Standards

In educational discourse, standards differ in their characteristics along at least four dimensions:

- the *type* of standard
- the *focus*; that is, the thing or event to which the standards are being applied
- the underlying characteristic or *construct*
- the *purpose* or use to which standards will be put.

These four dimensions affect the way in which standards are talked about and represented. Standards with different characteristics typically invoke different terminology, concepts and connotations. That is, once the type, focus, construct and purpose are settled, other characteristics follow. Different kinds of standards are (or should be) expressed differently, are related to a constellation of other concepts and carry hidden implications for educational practice and outcomes. These implications can be intended or unintended as well as desirable or undesirable. Consequently, choices need to be made. We should consider carefully whether some ways of talking about standards and some ways of representing them may not be benign, especially for individual students, and therefore whether alternative approaches would be more beneficial.

The following parts of this section discuss each of these four dimensions in turn. Each involves several categories and, in some cases, subcategories. It will be evident that there are some restrictions on relationships between the dimensions. That is, a choice of category on one dimension may exclude some choices on another dimension. It is best to consider this classification scheme as suggesting a series of questions to ask when someone refers to standards; that is, 'which of these types, focuses, constructs and purposes do you mean?'

Different Types

In a previous paper (Maxwell, 2002a), I argued that there are at least five different meanings of the term 'standards'. These will be referred to here as different types of standards (since I am extending the discussion here beyond that single dimension):

- standards as moral or ethical imperatives (what someone should do)
- standards as legal or regulatory requirements (what someone must do)
- standards as target benchmarks (expected practice or performance)
- standards as arbiters of quality (relative success or merit)
- standards as milestones (progressive or developmental targets).

The first three types are, respectively, desirable, necessary or appropriate; the last two are outcome levels. The first is usually expressed through guidelines or professional codes, the second through performance requirements that imply the possibility of failure (such as requirements for approving a program or awarding a certificate) and the third through statements detailing expected (or targeted or typical) outcomes.

The last two categories highlight two different ways of representing categorically different levels of learning, performance or achievement (see the section on constructs below for a discussion of these three terms): levels of merit (or quality) and levels of development (or progress). These are two sides of the same coin—both are concerned with identifying a range of ordered categories against which educational outcomes can be judged. An important difference is the time frame for referencing the performance: for merit standards, a set of comparative levels of merit (for example, on a task, course or program that is considered to be finished and done with); for developmental standards, sequential stages of possible improvement over time (requiring periodic re-assessment to determine current status). Some examples are discussed later in the third section of the chapter.

Different Focuses

Three important questions define focus: What facet of educational expectations are we focusing on? What units of analysis are we interested in? What is the scope of the assessment? A range of indicative possibilities for each of these can be summarised as follows:

- *facet*: educational content; or educational delivery; or educational outcomes
- *unit*: country; or system; or school; or program; or student
- *scope*: test, or task; or portfolio; or semester; or course; or certificate.

All three questions need to be considered in identifying the focus. For example, it is possible for the focus to be on the content of a course within a school (such as for school registration or approval to teach science). Alternatively, say, the focus might be on the individual student's performance on an assessment task. Yet again, the

focus might be on country performance on an international test. The permutations and combinations are clearly too many to consider them all in this chapter.

Because of the prevalence of content standards and performance standards in educational discourse, attention is given in the third section of this chapter to the distinction between them. Content standards typically apply to schools and systems (what they ought or must teach); performance standards typically apply to students (what and how well they have learned), although with a shift of focus performance standards could also apply to schools and systems. Of course, all outcomes are gauged by assessing students, but the way this is done can differ according to whether the focus is the individual student or a whole system—for example, a sample of students rather than a full census can be used for system monitoring, which practice has shown allows richer and more authentic assessment to be conducted (see, for example, the United States National Assessment of Educational Progress <http://nces.ed.gov/nationsreportcard/>).

Different Constructs

Another important question is what characteristic or construct is being assessed. The choice of constructs shapes how we represent and express any relevant standards.

Some contrasting constructs in this context include:

- learning versus performance
- development (time-extensive, assessing interim progress) versus achievement (time-limited, assessing degree of success)
- criterion-referencing versus norm-referencing
- quality (how well) versus quantity (how much).

Each of these warrants some discussion.

Learning Versus Performance

The contrast between learning and performance is a persistent one, stemming from the claim that learning itself is unobservable and that we must depend on observable performances (including, especially, those involving speaking and writing) to infer its existence. Strictly, of course, this is true. Nevertheless, dropping reference to learning entirely focuses attention only on the observables. These are merely indicative of the learning. They need to be referenced to the underlying dimensions of the student's learning, such as developing concepts and skills.[1] Keeping learning as the primary goal, with performance being indicative of it, situates the student's present performance in the context of their ongoing development as a learner. To see

[1] We can learn much from other areas of human endeavour. For example, in Olympic swimming, swim time against other competitors is the determinant of a standard. While this is a helpful reference for progress and the likelihood of being an Olympic swimmer, it provides no evidence to the trainer or swimmer about how to improve time until consideration of components (dive, turns, stroke style) is undertaken.

the student as a learner is to see the student as more than a performer of separate, isolated tasks. Standards that service learning may need to be represented differently from standards for performing a task. This idea connects with the next distinguishing achievement and development.

Development Versus Achievement

In common language, achievement is defined as: 1. something successfully accomplished, especially by hard work, ability or heroism; 2. successful completion (*Collins English Dictionary and Thesaurus*, 1993). An alternative term is attainment. The reference is to effort and striving as well as accomplishment—a journey completed and done well. The implications are all positive. Unfortunately, that is not how the term is used in education, where typically it refers to comparative performance; that is, 'how well did you do compared to other students?' When we talk of reporting achievement, it is not typically a description of the things that have been successfully accomplished but a rating or grading of performance on a task, semester or course; that is, reporting not what was done but how well it was done.

For example, the five grades reported on the Queensland Certificate of Education at the end of year 12 are called Levels of Achievement <http://www.qsa.qld.edu.au>. These grades are criterion-referenced, in the sense that they have pre-specified descriptions (requirements) for what constitutes achievement at each level and provide the benchmarks against which student achievement is referenced.[2] Although the intention is to judge each student's achievement against the requirements for each grade and not relative to other students' achievement, nevertheless it is expected that the grades will differentiate the range of student achievement across the state (generally it is expected that most students will not reach the top two levels, High Achievement and Very High Achievement). Therefore, for many students the message is negative (the lowest two levels, awarded to large numbers of students, carry the labels Limited Achievement and Very Limited Achievement—not exactly indicating successful accomplishment of anything—and even the middle category of Sound Achievement seems to damn with faint praise). Grades such as A–E are somewhat less assertive in their connotations of relative success and failure but the underlying conceptualisation is the same—students must be measured against a set of differentiated standards that are designed to discriminate the students as well.[3] While this might be appropriate for awarding certificates, diplomas, degrees and professional licences (satisfying a passing standard is often sufficient—for example

---

[2] This form of criterion referencing is referred to, in this instance, as standards referencing. This accords with defining the five grade levels (standards) as defined by (referenced by) particular achievement requirements (standards). This sounds tautological—the 'standard' is defined by a 'standard' but highlights the focus on the standards of achievement. Terminological debates and different practices surrounding the terms 'criterion' and 'standard' are beyond the scope of this chapter.

[3] Discriminate is the term used in educational measurement. It has a neutral meaning (tell apart), not a pejorative one (treat unfairly).

in awarding PhDs), it is unclear that grades are an appropriate way to register student progress at most stages of schooling, where a more descriptive approach to reporting progress (or achievement in the dictionary sense) would seem preferable.

Achievement reported through grades contrasts with charting development over time. Development suggests progress and elaboration over time. This could imply something that is 'natural', but it could also imply effort and striving. In education, it ought to reference both: the natural unfolding of human development needs to be coupled with experience and challenge that encourages and shapes that development. The time frame here is more extended; that is, point-in-time reporting is situated within a longer process of increasing strength and complexity of knowledge and proficiency. In that case, development can be represented by a continuum, with the student advancing progressively through the steps or stages along the continuum. Properly executed, progression of students at different rates along such a continuum can be accepted as normal and unexceptional. Examples of such developmental standards or progression targets are found in Judo rankings (*kyu* and *dan* ranks), English–language-proficiency scales (such as the International Second Language Proficiency Ratings), attainment targets for Key Stage Assessment and Reporting in England, and the six levels (and part levels) for the Victorian Essential Learning Standards (Years P–10).[4]

The contrast between graded achievement and developmental targets is discussed again in the third section of this chapter. It is a key issue in relation to how educational standards might best be represented.

Criterion Referencing Versus Norm Referencing

Another well-known distinction in assessment is between criterion referencing and norm referencing. The key idea in criterion referencing, in the form referred to here, is the specification of a number of ordered categories representing different levels of performance, preferably with each defined by explicit statements of the characteristics of each category and preferably also with exemplars to illustrate each category; assessment involves judgment of which category best fits the performance. The key idea in norm referencing is deliberate rank ordering along a scale, usually through aggregating scores on items; this may be followed by subdivision of the scale into a smaller number of levels defined by cut-scores, often by fitting an a priori distribution.

In practice, these distinctions are blurred. Thus, in criterion referencing, it is anticipated that the categories typically will capture a range of possible performances, since some degree of differentiation among students is expected; that is, there is an element of norm referencing underlying generation of the categories. Conversely, in norm referencing, each level can be described in terms of the characteristics typical of that level. An important difference is that criterion referencing

[4] Information on Judo ranks can be found at <http://www.judoinfo.com/obi.htm>, on the International Second Language Proficiency Ratings in Wylie and Ingram (1999), on Key Stage Assessment and Reporting at <http://www.qca.org.uk/>, and on the Victorian Essential Learning Standards at <http://www.vcaa.vic.edu.au/>.

establishes the categories or levels prior to assessment, whereas norm referencing does this afterwards. In addition, criterion-referenced categories serve as targets for student learning, whereas norm-referenced categories can only do so if they are carried over from one testing occasion to the next, in which case, for the second testing occasion they are predetermined and therefore criterion referenced.[5]

Quality Versus Quantity

Sometimes, a distinction is made between standards defined by quality (how well has the student performed?) and quantity (how much has been learned?). This is essentially an inappropriate distinction. The tradition of multiple-choice testing with a focus on right/wrong answers has not helped because better performance is equated to number of correct items. This can easily end up valuing more knowledge rather than more sophisticated and elaborated knowledge. As Shepard (2000) has pointed out, such practices are largely based on outmoded psychology. Current understandings of knowledge go much beyond recall and recognition (Pelligrino, Chudowski, & Glaser, 2001), and learners are better seen as charting a course from novice (the beginning learner) to expert (the proficient performer) (Bereiter & Scardamalier, 1993; Bransford, Brown, & Cocking, 2000a; Chi, Glaser, & Farr, 1988). Education needs to be concerned with understanding and using knowledge, including problem solving, creative endeavour and habits of mind (Costa, 2001; Costa & Kallick, 2000; Wiggins, 1998). As more is learned, it needs to be expressed and used in qualitatively different ways. Standards need to incorporate the qualitative progression from less complex to more complex knowledge and skill.

Different Purposes

Some possible purposes for standards include:

- setting targets for student learning
- showing students how they are progressing
- promoting consistency in judging achievement/progress
- setting requirements for qualification (certification)
- interpreting performances on tests
- setting benchmarks for system monitoring
- accountability for schools and systems

These purposes could be held in conjunction with each other. On the other hand, that could depend on choice of type, focus and construct. For example, as already

[5] There are, however, important differences between situations in which student performance is judged directly (classified) against a set of such standards (descriptive categories) and situations in which such standards are the basis for determining cut-scores on a continuous distribution (such as scores produced on a test). The language used to represent the standards could be similar but the processes for applying them are quite different, as too are the performances to which they refer. See, for example, Bennett (1998).

discussed, developmental standards could be more appropriate for showing students how they are progressing, while merit standards might be more appropriate for test results. In addition, accountability of schools and systems could be norm referenced but consistency in judging achievement depends on criterion referencing. Setting targets for student learning could be for individual students or systems and be linked to accountability or merely offered as a guideline (desirable but not mandatory). And so on.

Three of these purposes warrant special comment here.

Setting Requirements for Qualifications For qualifications, the critical standard is the minimum (or passing) standard required for receiving the relevant qualification (though this might be constituted of several standards, or requirements, in combination). This standard determines the qualification's social acceptability and credibility. Its function is sometimes to provide protection against professional incompetence (by only qualifying those with an acceptable level of capability) and sometimes simply to provide a 'tick of approval' (recognition for achievement). However, sometimes the relevant standard is determined by use of the qualification as a selection gateway, in which case the standard is set implicitly by the imposition of a quota rather than explicitly by a predetermined benchmark.

Setting Benchmarks for System Monitoring Another common purpose of standards is to establish benchmarks for system monitoring and/or accountability. For monitoring purposes, there are differences between setting benchmarks within a country or state linked to a specific curriculum (allowing teachers to use them formatively as targets for student learning) and setting a range of performance levels *ex post facto* on national or international testing programs (to allow richer comparison through descriptive performance levels rather than simply through scores).

Accountability for Schools and Systems Setting standards for accountability based on country or state testing programs is controversial and can have unfortunate side-effects. Sometimes, the target is unachievable, for example, the 100 per cent benchmark success target for literacy and numeracy benchmarks in Australia (<http://www.dest.gov.au/sectors/school_education/>) and the average yearly progress targets under No Child Left Behind (NCLB) in the United States (<http://www.ed.gov/nclb/>). The former was considered as a desirable goal (ought) but the latter was a required goal (must) so the consequences have been different. Failure to consider the consequences, such as narrowing of the curriculum, can be damaging.

This concludes the consideration of the different meanings of the term 'standards'. As previously suggested, I have suggested some questions that should be asked in any discussion of standards—questions concerning the intended type, focus, construct and purpose. Unless these questions are asked, it will often be the case that different participants in discussions about standards will have different implicit understandings of what is being discussed and will accordingly talk past each other.

The next section of this chapter considers some other aspects of the way we talk about standards and their implications for educational practice and student learning.

Setting Standards for Student Learning

Setting standards as targets for student learning is one possible purpose for having standards, as discussed in the previous section of this chapter. Two key ideas warrant further discussion: setting minimum standards and setting high standards.

Setting Minimum Standards

Setting minimum standards for student learning sounds like a good idea because it affirms an interest in and commitment to focused and purposeful learning. However, setting minimum standards implies the possibility that some may fail to reach them. How we handle such failure, whether we even refer to it as failure, is very important for the students concerned. While some educational commentators consider that the threat of failure is motivating, and it can be for some students, this is not so for all. In fact, it is bad psychology. The threat of failure can generate feelings of panic and inadequacy (Hodgson & Spours, 2005), often disrupting rather than supporting effective learning. Furthermore, failure itself can be accompanied by feelings of shame and rejection. The effects on students of being repeatedly classified as a failure (for some, over 24 semesters of schooling) can be catastrophic for the individual, and later for society through transformation of feelings of failure and inadequacy into anti-social behaviour. In the 21st century, we need to recast the educational language and practice so that such negative consequences are ameliorated.

For students, satisfying the minimum requirements for satisfactory completion of a stage of education is about 'permission to proceed' (that is, move on to some new endeavour to which the qualification provides access), 'permission to practice' (that is, admission to the profession or trade) or at least 'permission to claim the qualification' (that is, simply holding the certificate, diploma or degree as part of a curriculum vitae). In some situations (for example, in some countries for high school diplomas or university entrance examinations), students who fail in their attempt to gain a passing standard may have to redirect their energies and try something else. In other situations they may be able to try again. Also, in some countries, 'permission to proceed' is still relevant at points earlier than the end of secondary education, whether for access to the next stage of schooling or for streaming into different secondary schools or programs. Sometimes this involves a qualification (for example, in England, General Certificate of Secondary Education) and sometimes not (for example, primary school examinations as early as year 3 in some countries), but the effects of failure are similar—denial of access to further education or to particular programs.

The application of minimum standards where there is no selection or streaming is a futile practice unless failure is followed by some helpful action, such as remediation, or repeating the year or program, or redirection into some other activity. Some school systems, for example, in the United States, require such 'failing students' to repeat the year (but how many times before the student drops out?). Other school systems, for example, in Australia, value keeping the student age cohort together

(hence 'years' not 'grades'). Repeating a year rarely leads to better performance the second time around and in some cases to worse performance (as shown by Heubert & Hauser, 1999; Shepard & Smith, 1989). Typically, neither approach (repeating or progression) involves a tailored response to individual student needs. It is peculiar and damaging, therefore, to classify students as above or below an expected standard when there is no systemic way of properly managing students labelled as failed learners. There is an element of blaming the learner for this rather than blaming the inbuilt assumptions of school and curriculum structures. An aim for the 21st century should be to find a way of designing learning systems that are more personalised and adaptive.

A softer version of minimum standards is expected standards. How much softer depends on the force behind the word 'expected'. Leaving aside 'have to' (essentially the same as 'required'), expected can mean either 'ought to' (a moral or ethical imperative, but not one with serious consequences for failure, merely disappointment) or 'desired' (a learning target—what we would like to achieve if at all possible and therefore should work very hard to reach). Such expectations are typically framed for years (grades) or junctures (stages) and considered applicable to all students. This is again insensitive to individual student development and the diversity of student characteristics and capabilities in a classroom. Any such standards are typically directed at a level that some students will not reach within the designated time frame (such as before the end of year 7). Human development is too varied for that. It is futile to express such expectations if we know in advance that all students will not reach them (and, on the other side of the coin, that some students will exceed them by the proverbial mile). In the 21st century we should move to establish individualised learning targets that challenge students while respecting their own developmental possibilities (a *zone of proximal development* approach) (Chaiklin, 2003; Daniels, 1993). Maybe, with proper support, some students, rather than differing in ability, just need more time than others to attain a desired standard—something first suggested by John Carroll in the 1960s (see Carroll, 1963, 1989).

High Standards

Another part of the discourse on standards refers to 'high standards'. These are really 'high expectations'. For example, Wiggins (1991) says: 'A school has high standards when it has high and consistent expectations of all students in all courses' (p. 18). Also, Hill and Crévola (1999) talk about setting 'high and challenging standards that most students are expected to achieve' and suggest that 'low standards' are unacceptable ('zero tolerance of educational failure'). Is this reasonable?

Some research suggests that people respond to challenging or demanding expectations (the Pygmalion effect), though this is not necessarily so—there are interactions with various personal predispositions (Ng & Bahr, 2000). The assessment

literature carries many exhortations for teachers to set high standards for their students as a powerful strategy for increasing overall performance levels. The success of this strategy depends on continual encouragement and support, continual feedback that shows progress towards the target and personal belief that the target is achievable. Targets that are too high and believed by the student to be unreachable within the constraints of time and opportunities for learning are counterproductive, merely leading to frustration (Schunk, 1984). What is a high standard for one student may be too high for another student (beyond their zone of proximal development) or too low (undemanding and unchallenging) for another.

Another aspect of this stress on high standards is raising the ante with students pedagogically: 'by requiring students to work until standards are met, we teach students and teachers that work is not done until it is done right' (Wiggins, 1991, p. 22) which suggests a polishing of meta-cognitive skills, not just acquiring a lexicon of knowledge. There is also a value dimension operating here:

> When we speak of persons or institutions with standards—especially when modified by the word high—we mean they live by a set of mature, coherent, and consistently applied values evident in all their actions. (p. 20)
>
> Higher standards are not stiffer test-result quotas but a more vigorous commitment to intellectual values upheld consistently and daily in the face of entropy, fatalism, and the occasional desire on everyone's part to not give a damn. (p. 20)
>
> High standards are only to be found in completed tasks, products, and performances that require such intellectual virtues as craftsmanship, self-criticism, and persistence; when complex tasks are done consistently well, we easily and validly infer that the worker has high standards. (p. 21)

In fact, having high standards (expectations) in this sense is not just a matter of 'expecting a lot' and 'pushing the pace'. Expectations have to be realistic, that is, achievable with reasonable effort in the available time and prevailing circumstances. Even then, what may be realistically achievable in general or on average or by some is not necessarily achievable by everyone. The very concept of high standards in education has a norm-referenced underpinning. Standards reached easily by most people are ordinary, not high. There must be at least a sense that most people will struggle to achieve high standards. In other words, high implies contrast and comparison. However, the comparison need not be with other students in the same group; 'other people' can refer to other groups (other schools, other states or other countries) or other times (groups in previous years). Maybe, asking how high is high, in some objective sense, is a bit like asking how long is a piece of string.

Educational policy currently is awash with the need for 'high standards'. However, unreasonable expectations at the personal level, pursued inexorably, have dire consequences for student engagement and self-image. The resolution of this tension between group and individual expectations and progress remains an unresolved issue in educational theory and practice. Perhaps we will learn how to address this issue by the end of the 21st century—but only if we work on it.

Some Implementations of Standards

This section deals with three specific ways in which standards are characterised: content standards; performance standards (focusing on merit or proficiency standards); and developmental standards. These were chosen for their prominence and importance in educational discourse and practice. Examples are drawn from several countries around the world.

Content Standards

Sometimes, especially in the United States, 'standards' mean 'content standards'. These standards attempt to list the concepts and skills that should be the focus of teaching in schools and are usually organised by school subjects and by year/grade. In other places, they might be called a syllabus or a curriculum framework. Such standards could be considered as a moral or ethical imperative (type), specifying educational content for schools to use in framing their whole curriculum (focus), deal with what schools should teach (construct) and for target setting and accountability (purpose).

Content standards provide a 'road map' for schools and teachers, providing an overall structure of knowledge for each domain of knowledge and a framework for planning and delivering the curriculum. Their purpose is to ensure orderly progression and comprehensive coverage of important concepts and skills. Schools may, of course, repackage them to fit the way they wish to deliver the curriculum, including, for example, problem-based or interdisciplinary studies, though some content standards may constrain the extent to which they are able to do this.

Various United States agencies have developed standards for particular subject areas, for example, science standards by the National Research Council (1996), English, mathematics and science standards by New Standards (1997) and mathematics standards by the National Council of Teachers of Mathematics (2000). Every state has now developed its own standards for the core subjects. A recent review of these state standards by the Thomas B. Fordham Foundation (Finn, Julian, & Petrilli, 2006) claimed that only California, Indiana and Massachusetts had acceptable standards (their criteria being 'clear, rigorous, and right-headed [sic] about content'). For example, they considered that excellent English standards expect students to read and understand important literary genres, worthy science standards place the teaching of evolution at the centre of biology instruction and strong United States and world history standards are organised around a chronology of key events with an ample supply of fascinating and important individuals. Another highly critical review of state standards, specifically for science, is found in Wilson and Bertenthal (2001).

Content standards have the strength of giving teachers clear guidelines concerning the structural features of each subject and what may be appropriate for their students to learn at particular year levels. They provide scope and sequence for subjects.

However, content standards have the weakness of confusing what is being taught with what is being learned. They represent a 'backwards mapping' of each subject from the perspective of the 'expert' to fit the stages of schooling. Students do not necessarily learn according to the straightforward framework and sequence suggested by the content standards. The learning steps and sequences of the individual learner tend to be rather messier and unpredictable.[6]

Where content standards are packaged into tidy parcels of content to be taught to each year cohort, this inevitably means that some students will be left behind through not consolidating earlier material. Some other students may be bored by the lack of challenge unless given supplementary or accelerated learning opportunities. This phenomenon was well researched in the 'steering group' research of the 1960s and 1970s (Dahllöf, 1971; Kallos & Lundgren, 1979; Lundgren, 1972). That research showed that in teacher-centred classrooms where content is delivered in common to all students in a year cohort, the teacher typically and intuitively determines the pace of presentation (when to move on) by the rate of progress of the 'steering group'—students lying roughly between the 10th and the 25th percentiles of ranked achievement. In a curriculum organised and delivered by years (grades), the bottom 10 per cent of students are left progressively further behind because they cannot cope with the pace of new content. We have still not solved the problem within schools of how best to manage, both structurally and pedagogically, the substantial diversity within any group of students who typically are at different stages of development and learn in different ways and at different rates.

It is claimed sometimes that such (content) standards set uniform and high academic expectations for all students (Cohen, 1996; Darling-Hammond, 1997; Rowan, 1996; Sandholz, Ogawa, & Scribner, 2004). That is, the standards are what all students should know and be able to do (see earlier discussion). However, this makes them aspirational and idealistic. That is, it is unlikely that all students (strictly, perhaps, *any* student) will acquire all the knowledge and skills mentioned. Realistically, most students (perhaps all students) will acquire only some of the knowledge and skills (and retain some misconceptions and faulty skills as well).[7]

---

[6] Wilson and Bertenthal (2001) do call for content standards to reflect how students learn and develop understanding. While this can be broadly achieved by ensuring that more difficult concepts build on simpler ones, this still assumes that what works in general (for a typical student) will work for all.

[7] This is a necessary consequence of an expectation that all students will learn everything. Since clearly all students do not learn everything, the expectation that they will do so represents an unrealistic, aspirational target. The response to this is usually to recognise the spread of learning in some way, typically a range of performance standards. The top standard is the aspirational standard. Every other standard falls short and indicates some deficiencies of knowledge and skills. Students who just satisfy minimum requirements for a 'pass' standard (as opposed to the 'top' standard) are especially deficient and typically hold many misunderstandings, wrong ideas and inappropriate strategies. In systems that set 50 per cent success as a pass, students who just satisfy the pass requirement presumably do not know half of what was expected (or at any rate half of what was tested). Sequential content maps tend to assume that most students learn most of what was taught, which is clearly fallacious.

Performance Standards

A distinction is made typically between content standards and performance standards. For example, Stites (1999, no page numbering) says:

> Performance standards are specifications of 'how much' students should know and be able to do. Thus, while content standards shape what goes into a curriculum, performance standards set benchmarks—specified levels of achievement—that shape expectations for educational outcomes, provide a basis for measuring learning outcomes, and provide the criteria for imposing rewards and sanctions. Performance standards for mathematics, for example, specify the mathematical operations and concepts that should be mastered at each grade level as well as the types of assessments that should be used to measure that mastery.

The language in this quote may not be universal but it is instructive. It refers to benchmarks (type) for educational outcomes achieved by students (also implicitly teachers, schools and systems) at each grade level (focus), for measuring learning outcomes, quantity of knowledge and skills, level of achievement or mastery (constructs) and for accountability through rewards and sanctions (purpose).

Sometimes the line between content and performance standards is blurred, and performance standards become merely a more elaborated version of content standards. For example, The Georgia Department of Education provides the following explanation of its performance standards:

> Performance standards go into much greater depth than the content standards used in the previous curriculum. The performance standard incorporates the content standard, which simply tells the teacher what a student is expected to know (i.e., what concepts he or she is expected to master), and expands upon it by providing three additional items: suggested tasks, sample student work, and teacher commentary on that work. <http://www.georgiastandards.org/faqs.aspx>

Are these performance standards? Not really. Performance standards need to reference actual performance in some way. Where assessment is based on standardised tests, this might be represented by a cut-score (or several cut-scores if there are several standards). These can be arbitrarily defined (perhaps by an imposed distribution of levels or by natural breaks in the distribution of scores) and then given descriptive labels. Preferably, adopting a criterion-referenced approach, levels are defined by benchmark descriptions and cut-scores determined through a process of expert judgment (Cizek, 2001).

Another, now widespread, approach treats the benchmark descriptions as the performance standards against which assessors (typically teachers) make judgments of the level of achievement demonstrated by students. For example, in Canada, the British Columbia Ministry of Education defines performance standards in literacy and numeracy for each year/grade using generic labels that are elaborated by snapshot descriptors for each grade as well as by further elaborations for each 'aspect' (dimension) (see <http://www.bced.gov.bc.ca/perf_stands/>).

Interesting features of the British Columbia example are:

- The standards labels reference 'expectations' (not yet within expectations, meets minimal expectations, fully meets expectations, exceeds expectations): these labels signal a meaning for the standards beyond their indicating simply an ordered set of categories (range of proficiencies).

- The snapshot descriptors refer to the degree of completeness, familiarity and independence evidenced by the student.
- The descriptive elaborations of each standard reference specific and observable actions.
- The snapshot descriptors and descriptive elaborations bring in additional factors at higher levels.[8]

These performance standards are applicable to semester reporting by teachers of their students' achievement (and therefore could also be called achievement standards). They provide sufficient detail to identify performance characteristics that differentiate one level of achievement from another.

Sadler (1987) provided an influential analysis of these types of standards, in which he drew an important distinction between criteria (dimensions) and standards (levels). He also pointed out that standards defined by verbal descriptions are necessarily fuzzy (not sharply differentiated), indicative (not prescriptive or definitive), imprecise (because language is) and contextual (assuming familiarity with intended meanings and applications). Common interpretation and consistent use of such standards can be assisted by exemplars but may also need deliberate action such as assessor training and moderation (Maxwell, 2001, 2002b).

A typical way of representing such performance (achievement, merit or proficiency) standards is through a rubric (criteria-and-standards matrix). Rubrics are arbiters of quality (type), applicable to a variety of assessment artefacts (focus could be task, portfolio, semester, course, certificate). When developed by teachers for local application, their purpose is to make marking more objective, consistent and defensible, as well as to guide student learning (the latter by creating a language for discussing what distinguishes better performance from weaker performance). Rubrics have become a common feature of educational practice and satisfy a need to make explicit the basis on which judgments of performance quality (merit) are made.[9]

How explicit a rubric should be depends on the circumstances of its use. The general intention is to signal the performance characteristics in sufficient detail to support consistent judgment of the fit between performance and level. It is possible to frame the levels without connotations of failure if positive statements are made about the characteristics of each level. However, lower levels are often framed as deficient in some of the characteristics of higher levels, implying failure; also, a particular (expected) level is often designated as a satisfactory or passing level. Sometimes, an overall grade (level) is reported for each student, requiring a 'best fit' judgment that considers trade-offs between several dimensions (criteria); the specifics for each student are 'lost' and the grade description depicts only a 'typical' student. This may be adequate for certification. However, the specifics are important

[8] In other words, these are not in the form of a rubric (a fully crossed matrix of criteria by standards). New criteria emerge at higher levels of performance as the essence of differentiating higher levels from lower levels.

[9] One website <http://www.rcampus.com/indexrubric.cfm> provides a tool for developing rubrics and claims to have some 30 000 'ready to use' rubrics.

for feedback (formative purposes), where the detailed profile of performance on separate dimensions would be more useful.[10]

Developmental Standards

In England, although there is no explicit use of the term 'standards' in relation to the national curriculum and its assessment,[11] each national curriculum subject charts progress through nine levels (1–8 plus exceptional) along several attainment targets (strands) in each subject. The levels are represented through paragraph-length level descriptions (LDs) that summarise the characteristics of performance typical of each level. Progress against the levels is assessed at the end of each key stage (Stage 1: Year 2, Age 7; Stage 2: Year 6, Age 11; Stage 3: Year 9, Age 14). A holistic, on-balance judgment is made of which level best fits each student's performance. These levels represent standards in the sense that they involve milestones (type), individual student educational outcomes for junctures or stages (focus), development judged against specific criteria (constructs) and showing students how they are progressing in terms of a common national framework (purpose).

The complete attainment targets and levels are found on the Qualifications and Curriculum Authority website <http://curriculum.qca.org.uk>. The following extracts for English: Writing are illustrative:

> Level 2
> Pupils' writing communicates meaning in both narrative and non-narrative forms, using appropriate and interesting vocabulary, and showing some awareness of the reader. Ideas are developed in a sequence of sentences, sometimes demarcated by capital letters and full stops. Simple, monosyllabic words are usually spelt correctly, and where there are inaccuracies the alternative is phonetically plausible. In handwriting, letters are accurately formed and consistent in size.
>
> Level 3
> Pupils' writing is often organised, imaginative and clear. The main features of different forms of writing are used appropriately, beginning to be adapted to different readers. Sequences of sentences extend ideas logically and words are chosen for variety and interest. The basic grammatical structure of sentences is usually correct. Spelling is usually accurate, including that of common, polysyllabic words. Punctuation to mark sentences—full stops, capital letters and question marks—is used accurately. Handwriting is joined and legible.
>
> Level 4
> Pupils' writing in a range of forms is lively and thoughtful. Ideas are often sustained and developed in interesting ways and organised appropriately for the purpose of the reader.

---

[10] Comprehensive advice on designing rubrics is given by Wiggins (1998).

[11] There has been much debate in the United Kingdom about standards, whether they are being maintained from year to year for the General Certificate of School Education at Year 10 and the General Certificate of Education: Advanced Level at Year 12, but this is a different issue. See Aldrich (2000), Baird, Cresswell, and Newton (2000), Goldstein and Heath (2000) and Wolf (2000) for some background.

> Vocabulary choices are often adventurous and words are used for effect. Pupils are beginning to use grammatically complex sentences, extending meaning. Spelling, including that of polysyllabic words that conform to regular patterns, is generally accurate. Full stops, capital letters and question marks are used correctly, and pupils are beginning to use punctuation within the sentence. Handwriting style is fluent, joined and legible.

Sainsbury and Sizmur (1998), in their analysis of the complexities of the LDs, highlight the challenges caused by clustering several (somewhat disconnected) dimensions into each statement and also by needing to look outside the wording of the statements to professional understandings of the underlying constructs. Further, Hall and Harding (2002) found little evidence of the development of the communities of assessment practice needed to generate consistent interpretation and use of the LDs. Beyond these challenges there are clear advantages in reporting progress in age-independent (and year-independent) steps along a continuum. Green (2002) suggests several: the efficiency of a common set of benchmarks across all years; depiction of progress as movement along a continuum; focus on achievable progress rather than fixed ability; and 'natural' differentiation at each age or year level. To these could be added feed-forward opportunities, that is, higher levels as targets for learning (Sadler, 1989) and the motivating effects on students of experiencing growth and success rather than receiving the same grade year on year (Dweck, 1986).

Another realisation of the notion of developmental standards is the Primary Language Record (PLR) (Barrs, Ellis, Hester & Thomas, 1988) which has been highly influential (over 100 000 copies sold) in the United Kingdom, United States, Canada and Australia. Originally developed for multilingual inner London schools, but then expanded to fit the national curriculum, it is adaptive to cultural and linguistic diversity within common reporting frameworks. Five levels are defined for each of two age ranges (Scale 1: 6–8 years and Scale 2: 8–12 years) in both reading and writing.[12] Levels have labels (for example, beginning to fluent) and paragraph descriptions. Scale 1 charts progression from dependence to independence and Scale 2 from being inexperienced to experienced.[13] The PLR has been popular because of its emphasis on careful observation and documentation of student performance in authentic situations, charting their progress in a positive and supportive way and using this to plan next steps in learning.

The Australian state of Victoria offers another example of developmental standards. The roots of this approach go back to the attempt to create a national curriculum in the early 1990s. A key feature of this curriculum was levels of progression across the years of schooling (Willis & Kissane, 1997). Each Australian

---

[12] Details are available on the Research and Projects page of the Centre for Literacy in Primary Education website <http://www.clpe.co.uk/>. See also Falk (1998).

[13] There is implicit overlap between the scales but no natural transition—a single scale might work better. An exemplary single scale was developed by the Queensland Studies Authority (QSA) for the writing component of the Queensland Years 3, 5 and 7: Literacy and Numeracy Tests (QSA, 2007). This had four dimensions and 12 levels, for ease of use divided into sections typical of each year level.

state and territory soon decided to go its own way, with some erosion of the original ideas. Victoria remained committed to levels of progression for curriculum planning, assessment and reporting. In 2006 the state implemented the Victorian Essential Learnings Standards (VELS), which builds on and incorporates the previous Curriculum Standards Framework (CSF) <http://vels.cvaa.vic.edu.au/>.

VELS has three strands (Physical, Personal and Social Learning; Discipline-based Learning; and Interdisciplinary Learning), interrelated through what is characterised as a triple helix. Each strand has several domains, which are split further into several dimensions. For each domain there is a table of 'standards and progression points' that describes six developmental levels over the 11 years of compulsory schooling, together with three progression points between each level. The levels represent typical progress at 2-year intervals from end of Prep to end of year 10.

The term 'standards' is here used in three different ways: first, the knowledge and skills expected to be taught in each of the strands (content standards); second, the levels and progression points for assessing progress (development standards); and, third, the typical or targeted level for each year level (expected standards). To complicate this further, the Australian government imposed a national requirement in 2006 that all schools report student performance to parents each semester on an A–E scale (Commonwealth of Australia, 2005). Under VELS, Victoria maintains an expectation that schools will continue to assess the standard (level) and progression point reached by each student, with computerised conversion to an A–E grade appropriate for each year level (and representation of the levels in terms of their year of typical attainment). These characteristics of VELS are both visionary and realistic, adhering to the benefits of charting student progress developmentally but acceding to governmental and parental expectations. Whether this will be successful or confusing remains to be seen.[14]

Benefits of developmental standards include that they: provide a clear set of steps from novice to expert; emphasise language and expectation of progress; allow for student spurts and plateaus; and make evident to students the progress they have made. Difficulties include that: the progression of steps may not apply universally and levels typically cover several dimensions (with problems of best fit, as for performance standards). Challenges include: how to combine developmental levels with expected levels without reverting to a language of failure and how to develop school structures to support developmental progression of students better.

[14] Referents for A–E in Victoria are defined relative to the expected level for each year: well above, above, at, below, well below. Other Australian states and territories have adopted similar generic descriptors (for example, excellent, good, satisfactory, limited and poor) that offer crude comparative indicators (almost certainly inconsistently applied by different teachers and schools) but convey no information about what the student knows or can do.

Conclusion

This chapter provides an analysis of some different understandings and applications of standards in educational assessment. The intention has been first to provide a framework of different meanings and referents for standards, second to analyse the implications for educational practice of some prevalent ways of talking about and using standards and third to consider some contrasting implementations of standards and their strengths and limitations.

The framework of meanings and referents provided in the first section of the chapter suggests some ways of asking questions (about type, focus, construct and purpose) to clarify which of many possible meanings is intended in any discussion of standards. This can help to ensure that when people are talking about standards they are all talking about the same thing and not talking past each other because their focus is different.

The analysis in the second section of the chapter focuses on two prominent calls in current discourse on standards: setting minimum standards and high standards. This analysis suggests that there are some situations where setting minimum standards may be necessary and even beneficial, but others where they can be inappropriate and even damaging; similarly, there are some situations where expecting high standards may be desirable and motivating, but others where they can be unhelpful and even destructive. Further, minimum standards are written for the typical or average student, and high standards are by their nature not achievable by everyone, so both represent 'a bridge too far' for some students. Assessment practice and educational structures need to become more adaptive to personal circumstances and needs.

The third section of the chapter looks at some salient kinds of standards (content standards, performance standards, and developmental standards), examples of their implementation around the world, roles and uses, assumptions and connotations and strengths and weaknesses. Each was seen to have some virtues but to face challenges. Each should not be confused with the others. Some overall conclusions can be drawn.

Content standards. These can provide useful structuring of domains of knowledge, signalling important concepts and skills that schools should teach and students should learn. However, two types of adaptation are needed. First, content standards are not themselves the curriculum but inform its construction; schools need to devise their curriculum to fit local circumstances using the content standards as an input. Second, individual students do not necessarily learn at the standard pace and in the logical sequence laid out by the content standards; student learning is dependent on a variety of factors that disrupt any such 'assembly line' expectation. Rather, the general flow of student development contains lots of eddies where learning needs to revisit and consolidate before moving on. The notion of a spiral curriculum, with its constant revisiting and extension of central ideas and themes (Bruner, 1966), another idea from the 1960s, is worth revisiting as a way of resolving this lack of fit between a linearly sequential curriculum and the idiosyncrasies of human development.

Performance standards. In the form of merit, proficiency or achievement standards, these are the most widespread kind of standard. They are presented as levels,

represented by labels, descriptors and (sometimes) exemplars to indicate and instantiate their conceptual meanings. Performance standards are used summatively to report differences in quality of performance and formatively to acquaint students with differences between better and weaker performance. Handled well, performance standards can challenge students towards excellence. Handled badly, their latent competitive nature and potential for failure can destroy incentive. Working out a helpful and balanced role for performance standards is a challenge for the 21st century.

Developmental standards. One of the problems with performance standards is that they fail to make explicit how students change over time, especially where the same labels (for example, A–E) and sometimes also the same descriptors (when generic descriptors are used) are continually applied. Students can appropriate the label to describe themselves (a C-student; a failure). Developmental standards provide progressive labels and descriptions that indicate the unfolding journey of learning and the milestones passed. Examples are not widespread and those that exist are struggling to justify themselves. Yet, they have enormous potential as a replacement for or supplement to performance standards. More development is needed, and it may be best to start with those areas of learning that are naturally developmental, such as language learning. This is an important agenda for the 21st century.

One issue that keeps recurring is that of target setting. The setting of a blanket target for all students, whether it be for the content to be learned, the proficiency standard for satisfactory performance or the developmental level to be reached, *by a particular point in time*, fails to recognise and respect the diversity of background and circumstances, stage of development, existing knowledge and skill, personal characteristics and learning needs of individual students. This poses a dilemma: How to indicate desirable targets for learning and performance while respecting the individuality of the learner and providing positive feedback for progress made? This dilemma is especially pertinent in the compulsory years of schooling, when support and encouragement are so important and the conditions for life-long learning are being established.

Setting expected standards as general targets can be useful for defining what it would be good or desirable to achieve in various stages of learning, as a guide to curriculum implementation. However, this is likely to characterise a typical or 'average' student and therefore miss the mark for individual students. How these targets are represented, talked about and assessed is therefore important. Targets will function better if they are negotiated to fit the circumstances and if students have continuing opportunities to meet them over time. That is, flexibility is needed for schools to determine what the targets should be for each student. The notion of individual learning plans is one that should be applied to all students. In the secondary school, this should be broadened to encompass the notion of differentiated pathways (Grubb & Oakes, 2007).[15]

[15] Grubb and Oakes (2007) argue for schools as collaborative learning communities, for differentiated pathways based on multiple conceptions of standards and for stakeholder involvement in setting target standards.

Throughout this chapter, the focus has been on the welfare of the learner. Equity considerations require that every student be treated as an individual. A challenge for the 21st century is to work out how to reconcile representations of standards at a general systemic level with the idiosyncratic circumstances and needs of the individual student. In order to maintain a focus on the learner, we need ways of talking about and applying standards that reflect the 21st-century state of knowledge about learning. This knowledge is certain to keep growing and we need to be ready to absorb that knowledge into educational, and especially educational assessment, practice and recast that practice accordingly. There are three ideas that may prove productive lines of enquiry: personalised learning; brain research; and being more descriptive.

Personalising. A conclusion of this chapter is that an aim for the 21st century should be to find a way of designing learning programs and assessment systems that are more personalised and adaptive. Personalised learning has already captured considerable attention as a key concept in future delivery of educational services (see Keamy, Nicholas, Mahar & Herrick, 2007; OECD, 2006). This is a promising direction of development. Further consideration is needed of how notions of standards can fit with processes of personalised learning.

Brain research. At the beginning of the 21st century we are just in the infancy of neuroscience research on the brain. Such research will eventually revolutionise our understandings about how people learn. The implications for educational practice in the future are likely to be profound (OECD, 2007). Some scepticism is warranted for some current claims on the implications of brain research (McCandliss, 2002), but some cautious suggestions are already possible (Bransford, Brown & Cocking, 2000b; Jensen, 2005: Wolf, 2001). What is clear is that each brain, and therefore each student, is different and distinctive. This has implications for the tailoring of learning expectations, and consequently learning opportunities, to the individual student.

Being more descriptive. Both performance and developmental standards require a judgment of best fit to a standard that is represented by a description that typifies the standard but will not, in general, exactly describe the individual performance. This gives a broad overview of the performance in terms of the level reached in relation to other possible levels (and also, consequently, in relation to the performance of other students). In other words, it is a summary. If the level labels only are reported, then it is a very broad summary indeed, carrying no information about the characteristics of the performance. While this is useful for some purposes, such as certification and accountability, it is useless for others, such as providing feedback to assist further learning. Furthermore, greater emphasis on the personal advancement of students against tailored targets means attending more carefully and deliberately to the detail of each student's learning. Consequently, an important challenge for the future is developing ways of characterising and recording student achievement to keep better track of student learning and to make transparent to both teacher and student what next steps are needed for the student to make further progress. Technological advances may assist in doing this more easily, but successful implementation depends on clear thinking about how to set personalised targets for student learning. This is where the emphasis needs to be in assessment strategies of the future.

References

Aldrich, R. (2000). Educational standards in historical perspective, *Proceedings of the British Academy, 102*, 9–37. London: The British Academy. (Also in Goldstein & Heath, 2000). Retrieved June 19, 2008 from <http://www.proc.britac.ac.uk/cgi-bin/somsid.cgi?page=volumes/pba102&session=770884A>.

Baird, J., Cresswell, M. & Newton, P. (2000). Would the real gold standard please step forward? *Research Papers in Education, 15* (2), 213–229.

Barrs, M., Ellis, S., Hester, H. & Thomas, A. (1988), *The primary language record handbook for teachers*. London: Centre for Literacy in Primary Education. Retrieved June 19, 2008 from <http://www.clpe.co.uk/researchandprojects/research_06.html>.

Bennett, J. (1998). *Setting standards and applying them across different administrations of large-scale, high-stakes, curriculum-based public examinations*. Sydney: New South Wales Board of Studies. Retrieved June 18, 2008, from <www.boardofstudies.nsw.edu.au/archives/occasional_papers/occasionalp1_assess.htm>.

Bereiter, C. & Scardamalier, M. (1993). *Surpassing ourselves: An enquiry into the nature and implications of expertise*. Chicago: Open Court.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000a). How experts differ from novices. In *How people learn: Brain, mind, experience and school* (pp. 31–50). Washington, DC: National Academy Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000b). Mind and brain. In *How people learn: Brain, mind, experience and school* (pp. 114–127). Washington, DC: National Academy Press.

Bruner, J. S. (1966). *The process of education*. Cambridge, MA: Harvard University Press.

Carroll, J. B. (1963). A model of school learning, *Teachers College Record, 64* (8), 723–723.

Carroll, J. B. (1989). The Carroll model: A twenty-five year retrospective and prospective view, *Educational Researcher, 18* (1), 26–31.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller, S. (Eds.), *Vygotsky's educational theory and practice in cultural context*. Cambridge: Cambridge University Press.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, New Jersey: Laurence Erlbaum Associates.

Cizek, G. J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, D. K. (1996). Standards-based school reform: Policy, practice, and performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 99–127). Washington, DC: Brookings Institution,

Commonwealth of Australia (2005). *Schools Assistance (Learning Together—Achievement Through Choice and Opportunity) Regulations 2005*. Canberra: Federal Register of Legislative Instruments. Retrieved June 19, 2008, from <http://www.dest.gov.au/sectors/school_education/publications_resources/profiles/Schools_Assistance_Regulations_2005.htm>.

Costa, A. L. (Ed.) (2001). *Developing minds: A resource book for teaching thinking* (3rd ed.). Alexandria, Virginia: Association for Supervision and Curriculum Development.

Costa, A. L., & Kallick, B. (2000). *Habits of mind: A developmental series*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Dahllöf, U. (1971). *Ability grouping, content and process analysis*. New York: Teachers College Press.

Daniels, H. (Ed.) (1993). *Charting the agenda: Educational activity after Vygotsky*. London & New York: Routledge.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist* (special issue: psychological science and education), *41*(10), 1040–1048.

Falk, B. (1998). Supporting diverse learners with the Primary Language Record. In D. Allen (Ed.), *Assessing student learning: From grading to understanding*. New York and London: Teachers College Press.

Finn, C. E., Julian, L., & Petrilli, M. J. (2006). *The state of state standards*. Washington, DC: Thomas B. Fordham Foundation. Retrieved June 9, 2008, from <http://www.edexcellence.net/doc/StateofStateStandards2006.pdf>.

Goldstein, H. & Heath, A. (Eds.). (2000). *Educational standards: Proceedings of the British Academy 102*. Oxford: Oxford University Press. Also available at <http://www.proc.britac.ac.uk/cgi-bin/somsid.cgi?page=volumes/pba102&session=770884A> retrieved June 19, 2008.

Green, S. (2002). *Criterion referenced assessment as a guide to learning: The importance of progression and reliability*. A paper presented at the Association for the Study of Evaluation in Education in South Africa International Conference, Johannesburg. Retrieved June 18, 2008, from <http://www.cambridgeassessment.org.uk/ca/digitalAssets/113775_Criterion_Referenced_Assessment_as_a_Guide_to_Learning._The_.pdf>.

Grubb, W. N., & Oakes, J. (2007). *Restoring Value to the high school: the rhetoric and practice of higher standards*. Tempe, Arizona: Education Policy Research Unit & Education Policy Research Unit. Retrieved 9 June, 2008, from <http://epsl.asu.edu/epru/documents/EPSL-0710-242-EPRU.pdf>.

Hall, K., & Harding, A. (2002). Level descriptions and teacher assessments in England: Towards a community of practice, *Education Research, 44*(1), 1–16.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academies Press.

Hill, P. W., & Crévola, C. A. (1999). The role of standards in educational reform for the 21st century. In D. D. Marsh (Ed.), *ASCD Yearbook 1999: Preparing our schools for the 21st century* (pp. 117–142) Alexandria, Virginia: Association for Supervision and Curriculum Development.

Hodgson, A., & Spours, K. (2005). The learner experience of Curriculum 2000: Implications for the reform of 14–19 education in England, *Journal of Education Policy, 20*(1), 101–108.

Jensen, E. (2005). *Teaching with the brain in mind* (2nd ed.). Alexandria, Virginia: Association for Supervision and Curriculum Development.

Kallos, D., & Lundgren, U. P. (1979). *Curriculum as a pedagogical problem*. Lund: LiberLäromedel/Gleerup.

Keamy, R. K., Nicholas, H., Mahar, S., & Herrick, C. (2007). *Personalising education: From research to policy and practice*. Melbourne, Victoria: Department of Education and Early Childhood Development. Retrieved June 24, 2008, from <http://www.eduweb.vic.gov.au/edulibrary/public/publ/research/publ/personalising-education-report.pdf>.

Lundgren, U. P. (1972). *Frame factors and the teaching process*. Stockholm: Almquist & Wiksell.

Maxwell, G. S. (2001). *Moderation of assessments in vocational education and training*. Brisbane: Queensland Department of Employment and Training. Retrieved June 24, 2008, from <http://www.trainandemploy.qld.gov.au/resources/about_us/pdf/moderation_report.pdf>.

Maxwell, G. S. (2002a). *Are core learning outcomes standards?* Brisbane: Queensland Studies Authority (now Queensland Studies Authority). Retrieved June 10, 2008, from <http://www.qsa.qld.edu.au/downloads/publications/research_qscc_assess_report_1.pdf>.

Maxwell, G. S. (2002b). *Moderation of teacher judgments in student assessment*. Brisbane: Queensland School Curriculum Council (now Queensland Studies Authority). Retrieved June 24, 2008, from <http://www.qsa.qld.edu.au/downloads/publications/research_qscc_assess_report_2.pdf>.

McCandliss, B. (2002). Brain-based education in *Encyclopedia of Education*. Farmington Hills, Michigan: The Gale Group. Retrieved June 24, 2008, from <http://www.answers.com/topic/brain-based-education>.

National Council for Teachers of Mathematics (NCTM) (2000). *Principles and standards for school mathematics*. Reston, Virginia: NCTM.

National Research Council (NRC) (1996). *National Science Education Standards*. Washington, DC: National Academy Press. Retrieved June 19, 2008, from <http://www.nap.edu/html/nses/>.

New Standards (1997). *New standards performance standards: Elementary, middle school and high school*. Rochester. Washington, DC: National Center on Education and the Economy.

Ng, C. & Bahr, N. (2000). Knowledge structures and motivation to learn: Reciprocal effects. *Queensland Journal of Educational Research, 16* (1), 76–106. Interaction of knowledge and motivation. Retrieved June 19, 2008, from <www.iier.org.au/qjer/qjer16/ng.html>.

OECD (Organisation for Economic Co-operation and Development) (2006). *Personalising education*. Paris: OECD.

OECD (Organisation for Economic Co-operation and Development) (2007). *Understanding the brain: The birth of a learning science*. Paris: OECD.
dsfsdfsd

Pelligrino, J. W., Chudowski, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rowan, B. (1996). Standards as incentives for instructional reform. In S. H. Furhman & J. O'Day (Eds.), *Rewards and reform: Creating educational incentives that work*. San Francisco: Jossey Bass.

Queensland Studies Authority (QSA) (2007). Years 3, 5, and 7 Literacy: Writing. In *2007 Test Reporting Handbook for the Queensland Years 3, 5 and 7 Literacy and Numeracy Tests* (pp. 15–24). Brisbane: QSA. Retrieved June 19, 2008, from <http://www.qsa.qld.edu.au/downloads/assessment/3579_handbook_reporting_07.pdf>.

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

Sadler, D. R. (1989). Formative assessment and the design of instructional systems, *Instructional Science, 18*, 119–144.

Sainsbury, M., & Sizmur, S. (1998). Level descriptions in the National Curriculum: What kind of criterion referencing is this? *Oxford Review of Education, 24*(2), 181–93.

Sandholz, J. H., Ogawa, R. T., & Scribner, S. P. (2004). The standards gap. *Teachers College Record, 106*(6), 1177–1202.

Schunk, D. (1984). Enhancing self-efficacy and achievement through rewards and goals: Motivational and informational effects. *Journal of Educational Research, 78*, 29–34.

Shepard, L. A. (2000). The role of assessment in a learning culture, *Educational Researcher, 29*(7), 4–14.

Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention*. London: The Falmer Press.

Stites, R. (1999). A user's guide to standards-based educational reform: From theory to practice, *Focus on basics: Connecting research and practice*, vol. 3, issue C. Retrieved from <http://www.ncsall.net/?id=352>.

Wiggins, G. (1991). Standards not standardisation: Evoking quality student work, *Educational Leadership, 48*(5), 18–25.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve performance*. San Francisco: Jossey-Bass.

Willis, S., & Kissane, B. (1997). *Achieving outcome-based education: Premises, principles and implications for curriculum and assessment*. Deakin, ACT: Australian Curriculum Studies Association.

Wilson, M. R., & Bertenthal, M. W. (Eds.). (2001). *Systems for state science assessment*. Washington, DC: The National Academies Press.

Wolf, A. (2000). A comparative perspective on educational standards. *Proceedings of the British Academy, 102*, 9–37. London: The British Academy. (Also in Goldstein & Heath, 2000). Retrieved June 19, 2008, from <http://www.proc.britac.ac.uk/cgi-bin/somsid.cgi?page=volumes/pba102&session=770884A>.

Wolf, P. (2001). *Brain matters: Translating research into classroom practice*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Wylie, E., & Ingram, D. E. (1999). *International Second Language Proficiency Ratings (ISLPR): General proficiency version for English*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University.