


Author details:

Claire Wyatt-Smith
Claire Wyatt-Smith
Professor of Educational Assessment


This submission addresses the terms of reference relating to:

- i) ensuring assessment processes are supported by teachers, and
- ii) the ability of assessment processes to support valid and reliable judgments of student outcomes.

In preparing this submission I have drawn on my experience as a secondary school teacher and Head of Department (English), extended membership of the senior and P-10 English Syllabus Committees, contributions as State Panel Chair (English), and some three decades of engagement with research, policy and practice as an academic in the fields of assessment and literacy education. I also draw on my assessment and policy experience in several countries internationally. My work, including several large scale national competitive funded studies, has focused on teacher judgment, standards, moderation and evaluative frameworks. A limited set of relevant publications from the body of work including two forthcoming books on assessment are referenced below.

Introduction

The Queensland senior schooling system is both externally moderated and standards-referenced. Its current form represents a fifth era of a fully school-based assessment system that been implemented since 19970, followng the abolition of external examinations in Queensland. Given the generational distance between that period and the present, the history of the system is less well known today by the teaching workforce and indeed by the wider public, so a sketch of the various eras and related changes in school-based assessment may be of interest to the Inquiry and can be provided on request to the author of this submissionⁱ.

In brief, this history is of significance to this inquiry as it shows the clear move away from a system that relied on terminal examinations as the sole source of evidence for judging student achievement to a system reliant on explicit connections across i) the official or intended curriculum in syllabus documents, ii) School Work Programs designed and supported by teachers, iii) teacher-designed assessment tasks, each with accompanying statements of assessment criteria and standards, and iv) external moderation processes as part of the system's quality assurance checks and balances for both validity and reliability.

The centrepiece of the changes to the school-based assessment system in Queensland, throughout its iterations, has been student learning. There can be no doubt that, in principle, the approach taken to senior schooling in Queensland, has sought to put student learning and the learner at the centre. In principle, the system has well developed accountability measures and transparency, with School Work Programs being available for parent scrutiny, and each assessment piece, for both formative and summative purposes, accompanied by stated criteria and standards to inform students about expected features of quality. These features are also intended to inform teaching and learning practice, providing vital information to guide the students' efforts at self-assessment and improvement.

The approach to standards-referenced assessment that currently operates in the state is sufficiently distinctive from other forms of standards-referenced assessment internationally to have merited considerable attention from assessment experts and a range of scholars internationally with interests in learning improvement and accountability. Indeed, it is fair to say that the system has high credibility internationally, especially in terms of how it enables assessment in a high stakes environment that can lay claim to high validity and high reliability. The fact is that most assessment systems that are high stakes privilege reliability and necessarily rely on highly controlled conditions for bringing forth evidence of student achievement, most systems taking the traditional pencil and paper approach, with examinations centrally controlled. The achievement of the Queensland system is to put in place over some four decades a range of quality assurance checks and balances consistent with a high stakes assessment system, while at the same time placing teacher judgement at the centre. This mix of features is indeed the aspiration of many systems throughout the world facing the obvious challenges that flow from assessment measures that are overly reliant on examinations as the sole source of assessment evidence. The recently published OECD report makes this clear. Further, it is fair to say that all assessment systems require ongoing review and refinement, and this observation holds for the system in Queensland, given the generational distance mentioned above.

Claims to validity and reliability

There are three elements of the Queensland senior schooling assessment system that are linchpin in delivering high validity and reliability in a high stakes environment. These are:

- i) an explicit focus on and the constructive use of standards and their use in both learning and teaching and in determining grades
- ii) the requirement that formative assessment and summative assessment are connected in classroom practice, and
- iii) external moderation involving the use of stated standards as part of a complex system of quality assurance and system checks and balances for achieving comparability of grades, and in turn, fairness to students.

Element 1: an explicit focus on teacher decision-making in assessment and the constructive use of achievement standards

There is no doubt that the Queensland system of senior schooling assessment places stated standards as well as teacher decision-making and judgment at the centre of classroom practice. The teacher routinely makes decisions both about how to design instruction for diverse learners, and also the assessment evidence to be collected and interpreted for various purposes.

The potential of the system in operation is that the teachers have available to them in official syllabus documents explicit information about the expected features or characteristics of quality in the full range of achievement standards. In turn, the role of the teacher is to inform students about these features so that assessment expectations are known and used to enable students to know what they are aiming for.

Underpinning this central tenant of school-based assessment in senior schooling is that the standards are named and known by teachers and students. They can function as goals and assessment can be taken out of the black box. Of direct relevance here are the Exit Achievement Standards for judging student achievement on course completion, as well as the task specific standards and stated criteria that students are expected to receive before they commence work on assessable pieces. The key role of standards is therefore to inform teaching practice, student learning and importantly, improvement efforts.

Teacher judgment and standards are central to a long overdue focus on quality classroom assessment, with direct benefit to teachers and students' efforts to improve student learning and to bring forth valid assessment evidence. The potential of the approach to standards in the Queensland senior schooling system is realised when teachers design rigorous, intellectually challenging assessment tasks, and when teachers and students work together to develop student knowledge about the expectations or characteristics of quality. In the current system, teachers and students have ready access to stated achievement standards that operate at the level of the subject and that are then applied to the assessable tasks students are required to complete as part of their course of study. In relation to the focus on validity, this Inquiry would do well to consider how the potential of the current system is being realized in practice. This includes consideration of the rigour of the assessments that students are required to undertake and how they reflect 21st century thinking about the nature of knowledge in particular disciplines as well as the integration of other vital competences, such as those recognized by the OECD, for example.

The notion of 'fitness for purpose' (Gipps, 1994: 2–3) relates to knowing that assessment comes in a range of forms (such as examinations, rich tasks, projects, coursework, experiments, presentations and the like), as well as how it fulfils different purposes and aligns with different philosophies and learning theories. A fundamental question for teachers to consider in achieving better assessment for improved learning is: 'What is the purpose of this assessment I am designing?' From this stance,

examinations have continued relevance when designed for the purpose of seeing what students can produce under restricted conditions and on a limited set of unseen tasks. When a teacher wishes to obtain a richer perspective of student capabilities, then, typically, traditional pencil and paper examinations alone are insufficient.

Element 2: Formative assessment and summative assessment as complementary

Assessment is never simply about set activities, grades and certification. It is inevitably about teacher and student roles and relationships. In the Queensland system of standards-referenced school-based assessment the teacher has opportunity to undertake formative assessment in rich and diverse ways that not only connect to student learning in powerful ways, but that can enable student learning and improvement in ways not otherwise possible. Sadler's (1989, 1998) work on 'formative assessment' provides a model for a teaching-learning-assessment nexus that shows how improvement can follow when students are inducted into assessment knowledge and expertise. The induction is taken to extend to the teachers' knowledge of relevant criteria and standards, and more specifically, how they can function for improvement purposes. The current Queensland system has this potential.

The senior secondary teacher in the Queensland system is in a role to design formative assessments to build student capabilities, carrying forward prior learning, and preparing students for academic success on the set pieces on assessments for summative purposes. The validity of summative assessments that students undertake needs to be understood in two ways: first in terms of the construct validity of the instruments themselves – are they actually assessing what they claim to assess, and second, the opportunities for students to learn the knowledge, skills and capabilities that are being assessed.

A strength of the Queensland system, in principle, is that by treating formative assessment and summative assessment as complementary, students' learning needs can be diagnosed and addressed (formative assessment), and prior learning built on for success in summative pieces used for reporting student achievement.

There is no doubt that the Queensland model of senior schooling assessment calls for highly developed assessment capabilities on the part of teachers, as well as some new thinking. Within the scope of the current system there are opportunities for exploring how teachers can guide students to develop dispositions including creativity, flexibility, and resilience. To take up these opportunities there will be some element of confronting and naming those traditional assessment mindsets and identities that have constrained learning and that would seek to shut down assessments to those of a more limited scope.

There is no doubt that developments in technology will bring with them new ways of assessing and, in turn, these will require shifts away from traditional mindsets as a prerequisite for creative thinking or action. This is mentioned here in recognition of the radical changes facing the generation of teachers who are more likely to be comfortable with traditional approaches and print-bound ways of thinking, and the exciting opportunities that teachers and students can take up as co-learners in exploring how domain knowledge can be used, created, shared and assessed. From this vantage point, the Queensland approach to formative assessment and summative assessment, in which both are connected to standards, represents a sound basis for moving forward and further strengthening the system.

Element three: External moderation, teacher judgement and published standards

External moderation and the working of teacher panels are both integral in the quality assurance mechanisms of the Queensland system of senior schooling in its current formulation. I bring to this part of the discussion first hand experience of working as a teacher as a member of district moderation panels as well as my experience in the role as State Panel Chair, Senior English, albeit some years ago. I also draw on findings from several large-scale studies of teacher judgement involving the use of standards and the practice of social moderation. Three points are mentioned in relation to the concerns with validity and reliability.

First, the claim to validity in the Queensland system rests in part on the requirement that official syllabuses will be adhered to in the design and delivery of School Work Programs, including the assessment components. The syllabuses for all subjects are developed with direct input from the profession and other discipline experts. A feature of the system, however, is that while the parameters for assessment are set, the school has the authority to carry these into practice and implement them, taking account of their context. At the centre of the system is the understanding that assessment is necessarily contextualised practice. This is a highly regarded feature of the system that balances school autonomy and the need for common standards to be applied across contexts.

Second, moderation as practised in senior schooling is a process that enables teachers within a school and across schools to reach consistency of their judgments against a common, external set of characteristics of quality. In post-assessment moderation, where meetings occur after the assessment has been completed, moderation operates as part of a system's checks and balances for demonstrating accountability, especially in terms of comparability of grades. In this accountability context the focus is on how published features (assessment criteria) and levels (standards) of quality have been consistently applied within and across school contexts.

Central to external moderation as practised in the Queensland assessment system are teacher judgment, standards, and professional dialogue about how the requirements of the standards have been met in school submissions of folios chosen to illustrate the standards. It is worth mentioning that in Queensland, the system of externally moderated, standards-referenced assessment has been operating for some decades in the senior years of schooling as a mechanism within the system's overall suite of checks and balances for quality assured validity and reliability through to certification of student achievement in the final year of the two-year program of study.

The approach to moderation requires support from schools and teachers. Historically, this has been forthcoming with due recognition to the benefits of external moderation teacher meetings. Beyond the accountability checks, it is clear that moderation meetings function also as a form of professional development. This is of course not its primary aim, though my direct observation was that teachers benefitted from the opportunities to see School Work Programs from other contexts and in particular, to share in disciplined conversations about assessment quality, the application of standards and judgement practice.

The term 'comparability' does not apply directly to the processes that teachers rely on to arrive at a judgment. It is accepted that these processes are necessarily internal, drawing on a complex mix of explicit and tacit knowledge, including that derived from evaluative experiences over time. It is therefore understood that judgment processes will vary from assessor/teacher to assessor/ teacher and context to context. To emphasize then, comparability is the outcome of informed use of the stated standards.

Third, I turn attention to the use of marks. A key feature of the state's senior schooling assessment system, mentioned earlier, is that it is standards-referenced. The basis for the award of grades in a program of study is therefore the qualities of the work evident in the folio. Similarly, the quality of each piece in the folio is judged against a set of stated criteria and standards. Just as standards do not have a fixed meaning and can change over time, so too marks (numeric scores or letter grades) have no definite meaning. Meaning has to be given to a score of say 4 out of 10 or a D. Such quality annotations do not tell the person who has completed the assessment or the other party who receives the information anything about the quality per se of the piece. That is not to say however that marks cannot function as useful shorthand for recording a judgement of quality made in reference to standards in the first instance. The key point is that stated standards and related exemplars provide useful information about achievement that a mark, in and of itself, cannot communicate.

A comment about assessment futures and opportunities for change in the Queensland system

The Queensland system of high-stakes assessment, including the three elements outlined above, has maintained public confidence for several decades. The foundations are strong. There is a sense of generational distance, however, from today and the

period when the central tenants of the system were widely understood and the workforce commitment was strong, evidenced by waiting lists for teacher membership of district moderation panels. This statement recognises the commitment of the teaching profession to the current system and also how the gains in professional practice have been remarkable, and recognised to be so internationally. It is however to highlight how this level of commitment, and the assessment literacies that the system calls for, may well need investment. They will not necessarily sustain itself.

It may be time for teachers taking up first time employment in Queensland senior schooling to be given opportunities to be inducted into the principles and practices of the system. This could be done through online provision for beginning teachers or for teachers coming into the state's classrooms from other jurisdictions or other countries. Further work on this would be useful in ensuring teachers' assessment readiness for high quality assessment practices in senior schooling, especially as these relate to assessment task design, the formulation of assessment criteria and standards, and their use in judgement practices. Such induction could dispel, for example, the myths about the use of numbers in marking in the current system and address once and for all the longstanding distinction between objectivity and subjectivity in assessment. The fact is that the human brain is the primary instrument of decision-making, irrespective of whether assessment is in the form of teacher-designed assessments or external examinations.

The challenge for the system going forward is to build on the success of the past four decades and realise the potential of the system through targeted improvements. This challenge is not restricted, however, to government and the related authorities. It extends to the higher education sector, and in particular, initial and ongoing teacher education in Australia.

In concluding on a positive note: The assessment field has a rich field of studies undertaken in this country and internationally that show clearly how to leverage improved student learning outcomes. They show in part how clarity of expectations for teachers and students is an important feature of good practice, and the key role of the teacher in using classroom and system data to inform practice. There is no doubt that quality assessment is key to improving learning and teaching. Further, there is no doubt that summative achievement standards can and should be linked directly to teaching and learning. This has been an underpinning principle of the current Queensland system for some years. As the system has matured however, it may well be time to look to the nature and function of standards in different disciplines. The connections between discipline knowledge and standards both in classroom practice and judgement should be central to the Inquiry's deliberations concerning validity and reliability.

References

- Gipps, C.V. (1994) *Beyond Testing: Towards a Theory of Educational Assessment*. Falmer: London.
- Klenowski, V. and Wyatt-Smith, C. (forthcoming) *Assessment for Education: Standards, Judgement and Moderation*. London: Sage.
- Sadler, D.R. (1989) 'Formative assessment: Revisiting the territory', *Assessment in Education: Principles, Policy and Practice*, 5: 77–85.
- Sadler, R. (1998) 'Formative assessment: Revisiting the territory', *Assessment in Education: Principles, Policy and Practice*, 5(1): 77–84.
- Smith, C.M. (1995) *Teachers' reading practices in the secondary school writing classroom: A reappraisal of the nature and function of pre-specified assessment criteria*. Unpublished PhD thesis, The University of Queensland. [Smith now writing as Wyatt-Smith.]
- Wyatt-Smith, C.M., Klenowski, V. and Colbert, P. (Eds.) (forthcoming) *The Enabling Power of Assessment for Learning Improvement and Accountability*. Dordrecht: Springer.

Definitions

- Formative assessment*: Occurs when assessment, whether formal (e.g., testing) or informal (e.g., classroom questioning), is primarily intended for, and instrumental in, helping a student attain a higher level of performance. Assessment that occurs prior to summative assessment and whose purpose is partly to guide future learning for the student.
- Standards referencing*: The process of giving meaning to marks assigned to student work by referencing the image of the work to pre-determined standards of performance.
- Social moderation*: The process of validating teacher judgments of the standard of student work by having those teachers' judgments reviewed by their peers, internally (within the same school) and/or externally (from a different school).
- Standard (achievement standard)*: Fixed point along the criterion describing/representing qualitative (and discernible) differences in student performance. Standards are the referents that underlie judgment of success or level of merit in a performance. The teacher-assessor judges which one of several designated standards best represents the characteristics of a student's performance; that is, what label to attach to the performance or what category (such as A, B) to place it in.
- Validity*: The degree to which an assessment instrument measures what it purports to measure. Validity is a multifaceted concept, traditionally defined in many different ways, each one emphasising one particular aspect and named accordingly (e.g., construct, content, criterion, face, predictive).
- Reliability*: The degree to which the results of assessment are consistent, dependable, or repeatable.
- Summative assessment*: Occurs when assessment is designed to indicate the achievement status or level of performance attained by a student at the end of a course of study. It is geared towards reporting or certification. (There is no necessary distinction between the content of, or conditions for, formative and summative assessment.)

ⁱ This history was first written as part of my doctoral research into teacher use of standards in Senior English (Smith, 1995).