

International Association of Educational Assessment Conference
symposium paper – ref no. WC0023 Moderation: Principles, Policy
and Practice Presentation

Externally moderated school-based assessment in Queensland – How we know that it works

Peter Jordan, acting Deputy Director, Student Achievement Division, Queensland Studies Authority

Jo MacDonald, acting Principal, Senior School, Varsity College and State Review Panel Chair, Geography

Abstract

In 1970 the last external Junior (Year 10) examinations were held in Queensland followed by the last external Senior (Year 12) examinations in 1972. Replacing the use of external examinations, a system of internal assessment and moderation, involving a wider range of assessment techniques, was implemented. Initially, this system resulted in teachers assigning students' grades based on a 7 point norm-referenced rating scale. During the early eighties the system changed from norm-referenced to 'standards based' using five descriptive achievement levels externally moderated by a review panel system of 'teacher experts' for approving school work programs of study based on accredited state-wide syllabuses and verifying students' achievements for state-wide certification.

In recent years in Australia and elsewhere there has been an increasing focus on accountability in education. This has generated an intense interest in student results and the assessment regimes underpinning these. Under the current Federal Government, the establishment of a national Australian Curriculum Assessment and Reporting Authority (ACARA) to oversee and develop national curriculum appears at this stage to exclude any considerations involving a set of national external examinations in some subjects. If it did, then such a move would be at odds with the past 38 years of externally-moderated school based assessment in Queensland secondary school education.

All assessment systems face the issue of comparability of results that are reported on certificates of achievement. The Queensland Studies Authority identifies comparability of student results through a process of external moderation with respect to levels of achievement awarded by schools. Students who take the same Queensland Studies Authority subject in different schools and who achieve the same standard through assessment programmes based on a common syllabus, are awarded the same level of achievement. This paper will explore Queensland's senior secondary system of externally moderated school-based assessment with respect to the underpinning principles of reliability and validity in an analysis of how comparability of students' results is ensured.

Introduction – the past is the key to the present

To understand how we got to where we are today with externally moderated school-based assessment, we only have to look at the past. We would like to focus on the effects of the Radford Report (Radford, 1970) and then the Review

of School-Based Assessment or ROSBA (Scott et al, 1978) on shaping the current system of Queensland's externally moderated school-based assessment in senior secondary education.

From 1910 to 1972, sixty-three years of public external examinations set by the University of Queensland came to an end. The beginnings of school-based assessment in Queensland secondary schools can be traced back to the rumblings of collective discontent from students, teachers, parents and politicians regarding external examinations, particularly the Senior Physics examinations of 1966 and 1967. In 1966, students and teachers were highly critical of the Senior Physics examination which they regarded as unrealistic and extremely difficult. In 1967 more criticism resulted when 68% of candidates failed to register a pass on the examination (Clarke, 1987).

At about the same time schools were beginning to face the challenge of providing for larger numbers of senior secondary students who were not considering a university pathway for further studies but who were unable to exploit many alternatives in tertiary education because they had not been invented then! In 1968, the then Board of Senior Secondary School Studies commissioned Professor George Bassett, Professor of Education at the University of Queensland, to chair a committee to investigate the possibility of introducing a 'leaving certificate' for the growing number of senior secondary school students who did not proceed to tertiary education (Clarke, 1987). The Bassett Report (Bassett, 1968) concluded that senior external examinations did not adequately meet the needs of most post-secondary school students and that both those who entered tertiary institutions as well as those who left for work would benefit from the introduction of a leaving certificate concurrent with the existing Senior Certificate.

In 1969 the Board of Senior Secondary School Studies discussed the Bassett Report and recommended to the Queensland Minister for Education, the Hon. Alan Fletcher, that a committee be set up to consider the findings in the report. On 18 July 1969, a committee chaired by Dr William Radford, Director of the Australian Council for Educational Research, was commissioned by the Minister for Education "to review the system of public examinations for Queensland secondary school students and make recommendations for the assessment of students' achievements" (Clarke, 1987).

A new system of assessment

The report of the Radford Committee was presented to the Minister for Education in May, 1970. It recommended the abolition of both Junior and Senior external examinations to be replaced by a system of internal assessment and moderation to achieve comparability between schools. Copies of the Radford Report were circulated to teachers, universities and education sectors, the business sector and the public generally. Response to the proposal for abolition of external examinations was mixed but, on the whole, favourable (Clarke, 1987). In December 1970, the Minister for Education introduced a Bill in State Parliament to give effect to the Radford Report proposals. In support of the Bill, the Minister stated at the time:

"This is what is suggested for students - continuous assessment by the teacher.

This would be more reliable and penetrating. But the most serious objection to the external examination is that it does not test the extent to which the objectives of the syllabus in each subject have been met. Only limited areas in the syllabus can be examined in the Junior and Senior examinations.

Let us admit, too, that the build up of tension in examinations constitutes a great disadvantage to a child whose achievement over the years is to be assessed in two hours.

The external examinations also bring about a rigidity within syllabuses. Teachers, in fact, consult past papers more frequently than they consult the syllabuses.”

(Qld Parliamentary debates, vol 255, Dec 1970)

In January 1971 the new education regulations supporting school based assessment came into effect.

The Radford years: 1971 - 1980

The Radford Report was implemented by a reconstituted Board of Secondary School Studies from 1971 to 1979. In that time there were some difficulties experienced by teachers, students and their parents due to a lack of understanding of the spirit of the Radford proposals. In addition, parents and students expected some form of entrance requirement for tertiary study or employment while teachers were concerned with the need for inservice education and anticipated higher work loads. Public concern was leveled at the need for comparability between student results from different schools being demonstrated. This was meant to be achieved through the moderation process used at the time involving teacher representation from schools and the use of a common scaling test to establish an order of merit list for determining tertiary entrance.

In 1976 the Board of Secondary School Studies set up a committee to investigate the findings of two independent reports published in the same year which criticised the implementation of the Radford Report: the 'Campbell Report' and 'Schools Under Radford'. In 'Schools Under Radford', Fairbairn et al conducted a series of questionnaires and interviews which indicated that, while teachers generally favoured internal assessment, much dis-satisfaction stemmed from the moderation system. According to the report, the moderation system was seen by some teachers to be just as restrictive on teaching and assessment practices as external examinations ever were. The Campbell Report also echoed the sentiments of professional freedom in assessment practices being stifled at moderation meetings. The committee set up by the Board to investigate these findings was chaired by Professor Ted Scott, Dean of Education at James Cook University. Initially, Scott acknowledged that some of the early problems under Radford were gradually being attended to as teachers came to grips with internal assessment and moderation. However, a major change recommended by Scott in 1976 was a shift from norm-referenced assessment to a standards based approach to assessment and reporting. This meant that students would be assessed on standards that they achieved rather than how they compared with other students.

The ROSBA years: 1981 - present

The final report of the Scott Committee, 'A Review of School-Based Assessment'

(ROSBA) was published in 1978, approved by State Cabinet in 1979 and announced by the then Minister for Education, the Hon. Val Bird, to be phased in over a five year period beginning in 1981. That was over 28 years ago and it is essentially the system of externally-moderated school based assessment that we have here in Queensland today. One factor contributing to the success of the Queensland system is that it sends powerful messages to teachers, students, parents and the wider community about what really counts through five key elements:

1. There are guidelines that teachers/schools must use in planning [the syllabus].
2. There are formal plans for student learning and achievement that teachers/schools must make [work programs based on statewide syllabuses].
3. Evidence of student achievement must be produced [folios of student work].
4. This evidence must be assessed against the guidelines and plans [teacher judgment based on pre-set standards].
5. There is a process for validating teacher judgments of student achievement [social moderation].

In a keynote address given at an Australian, Curriculum, Assessment and Certification Authorities (ACACA) conference in Perth in 1999 by Dr Carol Myford from Educational Testing Services, she had this to say:

“When people ask me who is on the cutting edge, my first response is, ‘look down under’ ... On re-reading a 1985 account of Queensland’s externally moderated school-based assessment, I remember thinking how truly revolutionary it was in it’s scope. Upon my second reading and taking into consideration the political realities of the late 1990s ... I find it even more remarkable. My reaction to this program has moved up at least two notches on the excitement scale” (1999).

The findings of a recent national study in Australia indicate that despite different content and assessment methods used across the country:

- Student work across the country from different state and territory jurisdictions can be successfully compared against each other
- There are many common features within the student work at all levels (grades A-E)
- There is common agreement of the necessary level of skills and processes that a student should produce to be awarded the highest grade.

The exercise has shown that teachers can arrive at comparable assessment of student achievement, using a common scale, against agreed exit standards whatever the method of assessment (Australian Education Systems Officials Committee (AESOC) – *Senior Secondary Reporting*, July, 2007).

How do we evaluate the Qld system of school-based assessment?

Two important dimensions by which any high stakes assessment system should be judged are reliability and validity. Other dimensions such as community credibility and equity may also be used to make judgements on assessment systems; however, Matters (2006) noted that validity and reliability should be fundamental considerations in assessment policy and practice.

Reliability and comparability

The 1990 review of Tertiary Entrance chaired by Professor Viviani recommended that systemic and independent research be conducted to determine how comparable assessment outcomes were across schools in Queensland given the system of school based assessment. Since then research has been conducted on the Queensland school based assessment system in terms of inter-marker reliability and teacher perception of consensus based moderation.

Independent Research

One of the early studies into the reliability of the Queensland system was based on empirical research into how consistent the system was in the assigning of levels of achievement. This research into inter-marker reliability in the Queensland system was undertaken by Masters and McBryde (1994) and has become an important study in providing strong evidence of comparability in the Queensland secondary system.

In the Masters and McBryde (1994) study, a sample of 546 student assessment folios were rated independently by two markers under different sets of conditions. The results of the study revealed a very high level of comparability, with an inter-marker reliability index of .94.

Figure 1. First and Second ratings of folios plotted against each other (Model 3, all subjects pooled) (from Masters and McBryde, 1994)

Figure 1 shows a scattergraph of results from independent markers for Model 3 within the research. Lines have been drawn to indicate differences of plus or minus half a level of achievement (5 rungs) and plus or minus a full level of achievement (10 rungs). Masters and McBryde (1994) compared these results with a number of studies undertaken in other Australian states examining the comparability of results from external examination markers. In relation to the results of the Queensland study Masters and McBryde (1994) concluded that, "These levels of agreement are significantly higher than levels of inter-marker reliability typically reported from independent assessments of student work-including independent assessment of external examinations" (p.iv).

A more recent study that examined the effectiveness of consensus based moderation was undertaken by Klenowski (2006) as research into the processes used to validate teacher judgement of Rich Tasks, the trial project undertaken in primary and lower secondary schools in Queensland.

The results of this study determined that consensus moderation processes were perceived by teachers to be a reliable process because of the trust teachers placed in the professional judgement of other teachers and in the learning that came from viewing the work of students from other schools. The study also concluded that teachers recognised the importance of the quality assurance process of sampling that was undertaken at a central level to monitor the local moderation process.

Sadler (1993) noted that this level of professional accountability does not attract

much attention in systems with external examinations but the acceptance of research in the interests of “equity and public accountability” (p.11) should be welcomed in a professional environment. The evaluation undertaken by Klenowski (2006) revealed a highly professional attitude to the process of moderation and therefore the likelihood of more comparable outcomes. This high level of professional accountability is also evident in the annual quality assurance processes undertaken by the Queensland Studies Authority.

Queensland Studies Authority Research

In terms of comparability the main research undertaken by the Queensland Studies Authority involves the process of random sampling which commenced in 1994 and has been undertaken each year since then. Sadler (1993) identified a weakness in the Queensland moderation process whereby the stratified sample of student folios used in the moderation process was self-selected by the school. It was proposed that a set of student folios be nominated by the statutory authority, selected at random, independently reviewed and the results analysed as a measure of the comparability within the system. Random sampling has also facilitated longitudinal research on the review process with the accumulation of quantitative data on comparability now over fourteen years.

The number of folios examined each year varies depending on the subjects selected for sampling but in 2008, 2250 folios were independently examined. In some years over 3000 folios have been examined. The number of subjects, schools and student folios has varied over the fifteen years of random sampling; however, the large size of the sample and the systemically nominated manner of the selection of the folios ensures a high level of confidence in the results of the research process.

Figure 2. Comparison of percentage placed in same level of achievement (Random Sampling of Assessments in Authority Subjects 2008 Report, p.11)

As shown in Figure 2 (Random Sampling of Assessments in Authority Subjects 2008 Report, p.11), the first year of random sampling in 1994 resulted in only 79% of student folios considered by random sampling review panels to be appropriately placed in terms of the level of achievement; however, since then this has risen to at least 85% each year and was 93% in 2005. It should be noted that the levels of achievement assigned by the school also reflect the outcomes of advice provided in the local review process. This demonstrates a high level of

consistency in teacher judgments in terms of the five broad levels of achievement.

In addition, when the comparison of rung placement within the levels of achievement is examined it reveals a similarly high level of consistency of teacher judgment. In terms of the 2007 Year 12 cohort, 32% of all student folios were placed by random sampling panels on the same rung as the school and 56% were placed within 1 to 3 rungs of the school placement. Another 10% of student folios were placed by random sampling panels within 4 to 6 rungs of the school placement, with less than 2% placed more than 6 rungs from the school placement (Random Sampling of Assessments in Authority Subjects 2008 Report, p.10). On a fifty rung scale this demonstrates a very high level of comparability of teacher judgments and reinforces the findings of Masters and McBryde (1994).

Validity

While public debate often centres on the reliability of teacher judgements under different assessment regimes, there is usually less concern expressed on the validity of the assessment system. And though less research has been devoted to it, the structure of the Queensland system of school based assessment focuses strongly on the dimension of validity.

Traditional validity frameworks have assumed standardised forms of assessment such as a test and adopted a scientific approach to the analysis of the worth of the test. This approach promoted reliability potentially at the cost of validity (Moss, 1994). Such an approach to validity is more appropriate in systems where systemic testing dominates the assessment regime. Moss described an alternative approach known as a hermeneutic approach to validity research. This involved “holistic, integrative interpretations of a folio of student performances which privileged teacher judgements”.

Matters, Pitman and O'Brien (1995) undertook a study on validity and reliability in the Queensland system and drew on the work of Cronbach and Moss to develop what they termed “the Cronbach-Moss framework”. They described this structure as one where “validity has a tighter definition, and reliability has broader definition” (p.4). Further to this study, they investigated two validity perspectives from Cronbach (1988) as applied to the Queensland school-based assessment system. Through discussions with practicing teachers they examined validity in terms of a functional perspective. The researchers established strong evidence of authenticity of assessment in Queensland through links between syllabus objectives and assessment criteria and of the abundance of information used in making decisions regarding student performance (1998). Pitman, O'Brien and McCallow (1999) also argued that central to the construction of validity in Queensland was the concept that the value of assessment lies in its contribution to learning.

The need to consider the purpose of assessment and its value to learning has been an underlying theme in the work of Sadler (1998). He proposed that transparency in assessment practices can do much to promote life long learning in students. By disclosing criteria and standards, and ensuring that the decision

making underpinning teacher judgments was visible to students, he suggested that the locus of control for learning can move from the teacher to the student and can empower students in terms of their own learning. Sadler extended the notion by suggesting that students be made members of the “guild of professionals”.

The Qld senior system of externally moderated, school-based assessment: how we know that it works

Externally moderated school based assessment has been shown to have very high levels of inter-marker reliability over many years. Queensland’s externally moderated school based assessment system promotes greater ownership in teachers and builds capacity through increased professionalism and informed judgements of standards. However, the value of the system lies in its capacity to enhance the learning process and empower students to become “insiders” in the assessment process.

References

Australian Education Systems Officials Committee (AESOC) – *Senior Secondary Reporting*, July, 2007.

Bassett, G., *Report of the Committee set up to Consider and Make Recommendations to the Board of Senior Secondary School Studies Concerning the Introduction of a Leaving Certificate (Bassett Report)*, August, 1968.

Campbell, W. J. et al., *Some Consequences of the Radford Scheme for Schools, Teachers and Students in Queensland*, AACRE Report no. 7, Australian Government Publishing Service, Canberra, 1976.

Clarke, E., *Assessment in Queensland Secondary Schools: Two Decades of Change 1964 – 1983*, Policy and Information Services Branch, Department of Education, Queensland, May, 1987.

Cronbach, L. J., 'Five perspectives on validity argument', in H. Wainer (Ed.), *Test Validity*, Erlbaum, Hillsdale, NJ, 1988.

Fairbairn, K., McBryde, B. and Rigby, R., *Schools Under Radford*, Department of Education and the Board of Secondary School Studies, Queensland, 1976.

Fletcher, A., *Qld Parliamentary debates*, vol 255, Dec 1970.

Klenowski, V., *Evaluation of the effectiveness of the consensus-based standards validation process*. Townsville, Queensland: James Cook University, 2006.

Masters, G. N. & McBryde, B., *An investigation of the comparability of teachers' assessment of student folios*. Brisbane: Queensland Tertiary Entrance Procedures Authority, 1994.

Matters, G., Pitman J. & O'Brien, J., *An Exploration of validity and reliability that tolerates distinct privileges: Standardisation and contextualised judgements*. Montreal: Paper presented at 21st annual conference of the International Association for Educational Assessment, 1995.

Matters, G., Pitman J. & O'Brien, J., Validity and reliability in educational assessment and testing: A matter of judgement. *Queensland Journal of Educational Research*, 14, 57-88, 1998.

Myford, C. M., *Assessment for accountability vs. assessment to improve teaching and learning: Are they two different animals?* Paper presented at ACACA conference, Perth, 1999.

Pitman, J.A., O'Brien, J.E. & McCallow, J.E., *High-Quality Assessment: We are what we believe and do*. Bled, Slovenia: Paper presented at the 25th annual conference of the International Association for Educational Assessment, 1999.

Matters, G., *Assessment approaches in Queensland senior science syllabuses*. A report to the Queensland Studies Authority. Brisbane, 2006.

Moss, P. A., Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258, 1992.

Moss, P. A., Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12, 1994.

Queensland Studies Authority, *Random Sampling of Assessment in Authority Subjects 2008 Report*. Brisbane, 2008.

Radford, W., *Public Examinations for Queensland Secondary School Students (Radford Report)*, Queensland Department of Education, Brisbane, 1970.

Sadler, R., *Comparability in school-based assessment in Queensland secondary schools*. Brisbane: Queensland Tertiary Entrance Procedures Authority, 1993.

Sadler, R., *Letting students into the secret: Further steps in making criteria and standards work to improve learning*. Brisbane: Paper presented at the annual conference of State and District Panel Chairs, 1998.

Scott, E., *A Review of School Based Assessment in Queensland Secondary Schools*, Board of Secondary School Studies, April, 1978.

Viviani, N., *The Review of Tertiary Entrance in Queensland, 1990*. Report submitted to the Minister for Education, Brisbane, 1990.