

## **High-Quality Assessment: We are what we believe and do**

---

**A paper presented by John Pitman at the IAEA  
Conference, Bled, Slovenia, May 1999**

---

**J. A. Pitman, J. E. O'Brien, J. E. McCollow**



**Queensland  
Board of Senior Secondary School Studies**

This material is copyright. It may be copied freely for the use of schools in Queensland. It may not be reproduced for sale without express permission.

© Queensland Board of Senior Secondary School Studies 1999.

---

**National Library of Australia Cataloguing-in-Publication data:**

Pitman, John, (John, A.).

High quality assessment: We are what we believe and do : a paper presented by John Pitman at the IAEA conference, Bled, Slovenia, May 1999.

Bibliography.

ISBN 0 7242 7664 5.

1. Education - Australia - Evaluation. 2. Educational evaluation - Australia. 3. Educational tests and measurements - Australia. I. O'Brien, (John, E.). II. McCollow, (John, E.). III. International Association for Educational Assessment (IAEA) Conference, (1999 : Bled, Slovenia). IV. Title

---

Queensland Board of Senior Secondary School Studies  
Floor 7, 295 Ann Street, Brisbane  
PO Box 307 Spring Hill Qld 4004

Telephone: (07) 3864 0299  
Facsimile: (07) 3221 2553  
Email: office@qbssss.edu.au

DR1132

---

## Acknowledgments

We acknowledge and thank Gabrielle Matters for her thoughtful comments, and Paul Bennett for his wise editorial advice.

---

## Abstract

Assessment is an integral part of the learning process. It is most adequate, comprehensive and authentic when it occurs close to the learning environment. It should be based on clearly articulated criteria and standards of achievement. An assessment regime must be fair to all students, deliver value for money, encourage reflection, and provide mechanisms for improvement to be made. Drawing on the ‘Cronbach–Moss framework’ involving tighter notions of validity and broader notions of reliability, we propose a set of constructs, principles and values that underpin a high-quality assessment regime, and the structures and procedures that deliver a high-quality assessment system.

---

## Introduction

One of the conference themes invites us to consider: ‘who are we? where are we going?’. In this paper we examine issues related to this theme, but pose the questions a bit differently: ‘what do we believe in? what do we do?’ and, importantly: ‘what is the relationship between the two?’. We identify features of a high-quality assessment *regime*, describe the structures and procedures of a high-quality assessment *system*, and provide details of two key characteristics of externally moderated school-based assessment. We list some factors that shape the contexts in which assessment systems operate and describe briefly how one contextual demand, tertiary entrance, can be accommodated. In answering the conference questions we provide information about a specific assessment system while providing a general model for analysing and comparing assessment systems.

---

## I. Constructs

Constructs can be defined as the broad, general ideas that provide the ideological frame on which an assessment regime is developed. The framework we outline here draws on the work of Cronbach (e.g. 1988) and Moss (e.g. 1994). This paper expands on earlier papers (e.g. Matters, Pitman & O’Brien 1995), which described the features of the ‘Cronbach–Moss framework’<sup>1</sup>:

- tighter notions of validity
- broader notions of reliability.

Other possible constructs include quality, equity and social justice, effectiveness and efficiency, learning and motivation to learn. While not dismissive of the importance of these, we will argue that, by applying the Cronbach–Moss framework, these and others can be accommodated under one or both of the constructs, **validity** and **reliability**.

---

1. As noted in earlier papers, we do not claim that either Lee Cronbach or Pamela Moss would agree with or endorse the framework or its application.

Cronbach (1988 p. 5) identifies five perspectives on validity argument: the functional, political, operationalist, economic and explanatory. These categories overlap—questions about the utility of assessment, for example, can be said to derive from either a functional or an economic perspective, or both. We see the categories as conceptually sound and convenient rather than taxonomic.

At the centre of our construction of validity is the assertion that the worth of assessment lies fundamentally in its contribution to learning. We care about how well assessment fits with educational goals, and we pose questions about the relationship between the demands on and delivery of assessment, and about the adequacy of assessment methods, instruments and conditions. Furthermore, in contrast to some other constructions of validity, the construction described here invites us to consider (as ‘the bottom line’) the *consequences* of assessment both for individuals and for institutions. Relevant curriculum and assessment must meet a diversity of needs: opportunity for further learning, opportunity for work, opportunity to be part of society. In the Cronbach–Moss framework, questions about appropriateness as well as adequacy are raised. Ethical issues that require the invocation of social values—such as different kinds of cultural knowledge and experience—should be confronted, not avoided.

In this construction of validity, debates among teachers about the nature and meaning of assessment contribute to, rather than discredit, assessment validity. We believe in and, under the Cronbach–Moss framework, encourage validity inquiry that expects teachers to ‘cast about for facts [as] a route toward alternative interpretations and toward improved testing techniques’ (Cronbach 1988, p. 13) while treating as problematic the ‘epistemological principles used to warrant validity conclusions’ (Moss 1992, p. 229). The approach entails consultation and negotiation amongst teachers. It also entails seeking the views of students about the content and timing of assessment and fostering an awareness in students of the decision-making process. We do not expect unanimity of beliefs or decisions. We do expect that each student learns how assessment decisions are made, and how they affect his or her interests, and that the student feels that the decision-making process is responsive to inquiry.

This construction of validity is, we would argue, a highly pragmatic one—grounded in the curriculum practices of teachers and cognisant of the consequences of assessment for students. Validity as a construct should, we believe, incorporate consideration of practical issues, such as meeting timelines and working within resource constraints.

Reliability is generally seen as a necessary but insufficient condition for validity. It is also seen as an important guarantor of fairness, safeguarding against the possibilities of subjectivity and bias. The problem is that, in the psychometric paradigm, defining reliability as a measure involving the calculation of differences between independent observations entails a significant degree of standardisation. A major tension thus arises between privileging standardisation (under the rubric of reliability) and calling for assessment to be more ‘authentic’; that is, standardisation is at odds with a notion of assessment arising naturally from the learning situation (which pertains to its validity).

For some assessment regimes, reliability’s perceived status as a precondition for validity has resulted in this tension being resolved in ways that elevate reliability concerns over validity concerns. (The proof is the degree to which standardisation is a hallmark of many regimes.) We, on the other hand, remind ourselves and others that without validity, reliability is trivial.

Given validity, we highly value reliability. We want to broaden, not obliterate, the latter<sup>2</sup>. What we argue for is not the overthrow of the psychometric paradigm, with its emphasis on measurement and standardisation, but that it should be complemented by, and where appropriate, give way to, a hermeneutic approach that emphasises the exercise of contextualised judgments based on firm evidence. Assessment is seen as a decision-making, not a measurement, process. This approach stands in its own right as legitimate. It also provides a productive way of questioning the epistemology of validity under the psychometric paradigm.

Psychometric processes aggregate results or judgments and the aggregates are compared with standards. Hermeneutic processes expand the role of judgments to involve integrative interpretations of standards of work based on all relevant evidence. The latter countenances the use of contextualised teacher judgments in a climate where ‘inconsistency in performance across tasks ... becomes a puzzle to be solved’ (Moss 1994, p. 8)—not, as statistical models would have it, disruptions to be smoothed—and a critical dialogue among stakeholders is encouraged. The outcomes of different methodologies—psychometric and hermeneutic—can be the same, but it is not necessarily so. Neither methodology is always appropriate, but thorough consideration of the merits and limitations of each (rather than the dogmatic defence of either) will lead to better assessment outcomes.

A broader conception of reliability can be said, therefore, to have three ‘layers’ (Moss 1994):

- privileging contextualised judgments
- generalising across tasks
- generalising across readers.

High-quality assessment will be valid *and* reliable.

---

## II. Principles and values

In this section, we describe the principles and values that inform, clarify and extend the constructs of validity and reliability.

Broadfoot (1992) suggests that the first task in designing an assessment system is to ask what is the goal. While acknowledging that there can be many goals of assessment systems (e.g. certification, selection) we share Broadfoot’s view that an assessment system should promote learning, and we believe this is fundamental. Without a highly shared awareness of this goal, an assessment system is deprived of value.

Assessment should be based on clearly articulated criteria and standards of achievement<sup>3</sup> which arise from syllabuses<sup>4</sup>, inform programs of teaching and learning, and shape assessment instruments. Assessment should be a natural part of the learning process, fostering the development knowledge and skills. Since the taught curriculum is the basis upon which assessment is built, it

---

2. Thus, our position can be distinguished from that of Jessup (1989, p. 193), who argues—specifically in relation to vocational competency-based assessment—that ‘we drop the concept of reliability ... and concentrate our energies on maximising validity’.

3. We use the phrase *level of achievement* to mean the result, derived from a combination of performances of a student, in a course of study.

4. We use the term *syllabus* for a document outlining the body of knowledge to be taught, the thinking and doing skills to be developed, and the way these are to be assessed.

follows easily that assessment should take place as close as possible to learning; that is, the locus of assessment should be the site of learning. In every area of life, except perhaps education itself, such a proposition would be accepted as axiomatic.

Because teachers are fully knowledgeable of the contexts in which learning occurs, assessment should be developed, administered and evaluated by the teachers who develop and implement the curriculum, and work daily with students.

To be adequate, comprehensive and authentic, assessment should consist of tasks which are representative of the 'knowledge, skills and strategies needed for the activity or domain being tested' (Fredericksen & Collins 1989, p. 30). Adequate assessment will provide evidence that student performances are consistently judged in accordance with prescribed outcomes. Comprehensive assessment encompasses the range of knowledge, skills and strategies across the domain being tested, and this should be reflected in a range and balance of assessment methods, instruments and conditions. Authentic assessment involves students in using relevant and useful knowledge, thinking and practical skills.

Opportunities for students to achieve are enhanced when they have access to the criteria used to make judgments about achievement (see Pitman & Dudley 1985, p. 17) and when they receive regular feedback on the standard of their work. Integrating assessment and learning offers students greater control of learning and its outcomes by helping them understand the desirable goals of learning and to self-assess progress toward these goals.

An assessment system must be fair to all students. Fairness involves providing information to students that allows them to be aware of the consequences of assessment. It should also include ongoing analysis to identify and avoid systematic bias against groups of students. System-related barriers to student achievement (for example, procedural barriers) can be removed by making alternative arrangements based on special consideration of individual circumstances.

Results in particular subjects<sup>5</sup> must be comparable not just within schools but across schools.

An assessment system should encourage reflection on the quality of assessment and provide mechanisms by which reflection about assessment methods, materials and results can be translated into real developments and improvements in assessment practice. Opportunities should be provided to develop and explore shared understandings on crucial questions such as: 'what should the standard be?' and 'what constitutes adequate evidence that the standard has been met?'

The decision-making processes of an assessment system should be transparent to students, teachers and the broader school community. Assessment procedures should involve consultation with stakeholders and opportunities for disputation, negotiation and resolution of differences.

---

5. We use the term *subject* to mean an area of study within a discipline. Physics and Chemistry, for example, are two subjects of the science discipline. Accounting is a subject of the commerce discipline. A senior secondary school curriculum may comprise fifty or more subjects. A student would generally study five or six subjects.

From the perspective of a student, a key economic feature of assessment is the utility of the knowledge and skills assessed and of any credentials associated with the assessment. From a system perspective, assessment should be cost effective; that is, it should deliver value for money.

Accountability should be infused into, rather than imposed on top of, an assessment regime and system. With infusion, accountability is devolved, in a way not possible with imposition, to the many levels (classroom, school, district, system, regime) and dimensions (fairness, comparability of results, quality, cost effectiveness) of assessment.

---

### III. Structures and procedures

For the purposes of this paper, we see constructs, principles and values as constituting an assessment *regime*. We turn now to the structures and procedures which operationalise the constructs and translate an assessment regime into an assessment *system*. In this section of the paper we describe in general terms one assessment system: externally moderated school-based assessment. Such a system operates in Queensland, Australia and has, in our view, been successful for more than twenty-five years in operationalising the constructs<sup>6</sup>. In section IV, we draw more specifically on the Queensland experience.

The assessment system we outline here reflects what Gipps (1994, p. 1) describes as the ‘paradigm shift ... from a testing and examination culture to an assessment culture’. Development of the latter culture is characterised by:

- assessment techniques and instruments that support teaching and learning, and foster wider knowledge and higher-order thinking
- recognition of the limitations of assessment
- efforts to provide assessment data for certification, accountability and other legitimate purposes in ways that do not damage teaching and learning.

In an externally moderated school-based assessment system, schools take responsibility for implementing assessment but participate in procedures through which teaching and assessment programs are endorsed. They agree to the outcomes of these and other procedures for the review of standards of student work and verification of results. Schools also agree to make teachers available to participate in these procedures, which are undertaken at meetings of panels of local teachers (review panels).

The curriculum and assessment authority is responsible for quality control of assessment outcomes and for establishing and maintaining community acceptance of a certificate issued at the completion of senior secondary schooling. It is not a teacher-employing authority and so does not have a teacher supervisory role—this remains with principals of schools on behalf of the employing organisation.

Critical discussion of assessment tasks with colleagues is part of internal (within-school) moderation. School administrators have supervisory and support roles involving review of assessment methods, materials and results. Together with teachers they have a responsibility to the school community to

---

6. School-based assessment was introduced in Queensland in 1972. It changed from norm-based to criteria-based assessment in 1981.

provide maximum opportunities for each student to learn and achieve. Review panels, supported by officers of the authority, have a crucial role as external (to the school) moderators. Review panels provide advice to schools about the suitability of assessment tasks and outcomes proposed by schools.

Thus, the system provides a structure within which school and teacher decisions about programs of study and assessment, and the actual assessment of student work itself, are verified. It involves a partnership between schools and a systemic curriculum and assessment authority. The shared goal of the authority and schools, and the fundamental reason for the partnership, is the accurate, coherent, comparable and valued certification of results.

Assessment is subject-based in that, for each subject, there is a syllabus developed by an advisory committee of teachers<sup>7</sup>, which outlines what needs to be taught and how it is to be assessed. A syllabus is a framework document, rather than a set of rules or specifications. Schools write work programs outlining a course<sup>8</sup> based on a school-focused interpretation of each syllabus, taking account of such things as a specific school's environment, expertise of its teachers and local community expectations. Each syllabus is organised in terms of a rationale, global aims, objectives, organisation of subject matter, learning experiences and assessment strategies. The syllabus conveys a set of values, and clarifies what teachers of a subject collectively believe to be the valued outcomes for students. Syllabuses also play the important role of defining the criteria and standards used for assessment.

Assessment is continuous in that it occurs over a period of time rather than on a single occasion. Continuous assessment is a balance between the undesirable extremes of incessant (e.g. daily) and quantum (e.g. annual or semesterised) assessment. Syllabuses provide guidelines for the maximum and minimum number of assessment scripts needed to form a student portfolio. Schools design programs incorporating a range and balance of assessment tasks that build, for each student, into a portfolio. Scripts in the portfolio may be updated so that the portfolio at any time contains the 'fullest and latest'<sup>9</sup> information about student progress and achievement in the subject. This adds robustness to teacher judgments and allows for different rates of student progress to be accommodated.

School-based *formative* assessment, incorporating features of continuous assessment, is, of course, common to many assessment regimes. In this system, however, summative assessment is also school-based. Black and Wiliam (1998, pp. 17–19) point out some of the inadequacies that can arise in teaching, learning and assessment when classroom teachers have only a (putatively) formative assessment role and assessment is dominated by external summative testing. A major tendency in such systems was for teachers to use the external tests as models for their own assessment, undermining its formative role, or to relegate formative assessment to assessment of only attitudes and behaviour, seriously devaluing high-order thinking.

---

7. Includes teachers from secondary schools and universities.

8. The teaching, learning and assessment program in a subject.

9. A phrase deliberately invoked to help carry the message to teachers that the sample of assessment scripts in a student portfolio should be representative of the complete (full) course while allowing that later scripts may be more representative of student achievement. A little more detail about this is provided in section IV.



In the Cronbach–Moss framework a broader notion of reliability countenances contextualised judgments in the assessment of student work. Sadler (1986b, p. 6) argues that teacher judgments ‘lie at the heart of good teaching and good assessment’. Judgments about student achievement can be validated by other teachers as members of the *guild* of professionals. It is not expected that consensus be reached on every single aspect of work in a student’s folio, but a complex on-balance judgment of achievement can be made by teachers and agreed on by review panels.

School-based assessment has been criticised as susceptible to the effects of subjective teacher judgments, leading, for example, to grade inflation (see Elley undated). Sadler identifies three conditions under which teacher judgment can produce comparable, not inflated, decisions about results: time to make the judgments and to reflect on them, internal moderation, and external moderation. Schools have a responsibility to meet the first two conditions; the curriculum and assessment authority, through the work of review panels, has a responsibility to meet the third. Where these conditions are met, ‘there is unequivocal evidence that teachers ... make consistently reliable and valid appraisals’ (Sadler 1986b, p. 6). The research of Masters and McBryde (1994) confirms that review panels are successful in delivering comparability of results and avoiding grade inflation.

In implementing a program of continuous assessment, teachers need to be careful to provide sufficient formative assessments to allow students to develop response techniques for the range of assessment instruments and conditions that will be applied. When scheduling assessment tasks teachers must be aware of the stage and rate of development of students to help ensure that there has been adequate time for students to learn sufficient subject content, so that assessment of understanding and application is grounded in that knowledge.

While continuous assessment makes demands on teachers, it also allows teachers the flexibility to meet them. By spreading assessment decisions over both time and tasks, not only is the evidence used to support judgments increased, so too are the opportunities for reflection on those judgments. Teachers can divide assessment into suitably timed and sized parts. For example, teachers can, from time to time, separate out routine assessment of knowledge and techniques and non-routine assessment of higher-order thinking skills. In this way *time allowed* is less likely to be a covert criterion of assessment. Furthermore, the timeframe involved in continuous assessment is an important aspect of putting together student portfolios and providing time for teachers to reflect, consult and negotiate with review panels about the assessment and its outcomes.

Continuous assessment also provides, in a way that terminal assessment cannot, both motivation and opportunities for students to reflect on their work, develop strategies for improvement, and demonstrate improvement.

Nevertheless, we do not want to understate the difficulties involved for teachers in meeting the challenges posed by continuous assessment. Continuous assessment makes time management demands on teachers and students, it can lead to over-assessment, and the tensions that arise between the formative and summative purposes of assessment need to be carefully thought through and managed. While it can be argued that combining the formative and summative functions of formal assessment leads to a reduction in formal testing and more time for teaching and learning, a high-quality assessment system needs to have

strategies in place to avoid over-assessment. The development of explicit criteria and standards, the involvement of review panels of teachers in accreditation and review, and the use of student portfolio records provide ways of overcoming these potential pitfalls.

The use of criteria and standards<sup>10</sup> involves identification of the important characteristics or properties for assessment, description of performance benchmarks in each characteristic, and description of ‘tradeoffs’—allowable combinations of performances—associated with each level of achievement. Each syllabus has a criteria and standards schema describing, on a continuum, benchmark performances in each criterion. This stands in contrast to norm-based assessment where the emphasis is on relating the achievement of a particular student to the achievements of other students.

As an integral part of devising fair student assessment tasks, teachers write task-specific criteria, which are declared in advance of implementation of the task and used to assess student work. Criteria sheets allow students to identify appropriate ways to demonstrate knowledge and skills and guide teachers in making decisions about standards of student work. Work programs are available to the school community so that students and parents can be aware of the nature of the assessment program.

Considerable emphasis is placed on developing criteria and standards that are explicit, well-articulated, well-understood, and easily internalised and applied by teachers. Criteria and standards facilitate clear thinking and communication amongst teachers, enhance the comparability and portability of results, and help students (and their parents) understand assessment decisions. In addition to providing a structure that considerably enhances the quality of feedback provided to students about their performances, the prespecification of criteria and standards constitutes a form of ‘feedforward’ (Findlay 1987, p. 6). In the context of school-based assessment, criteria and standards serve the important function of reducing the possibility that individual teachers’ decisions will be (or will be seen to be) arbitrary.

The system we describe empowers teachers to implement high-stakes assessment and stimulates professional conduct. The authority can actually promote change and variety, rather than uniformity, in school assessment practices. Teachers in schools are encouraged to focus on the links between assessment and learning. They have freedom to determine the:

- learning and assessment experiences of students
- time allocated to assessment
- frequency and timing of assessment
- nature of reporting to parents and feedback to students.

This freedom is not, of course, unconstrained. Some freedom is bounded by syllabuses. Some freedom is bounded by review panels. And accompanying this freedom are significant responsibilities. Teachers have a responsibility to ensure that assessment is of high quality. With the support of colleagues in the school and on review panels, and of administrators and officers of the curriculum and assessment authority, teachers are responsible for devising and implementing assessment that:

---

10. Following Sadler (1986a, p. 5), we define *criterion* as the ‘property, dimension, or characteristic by which something is judged or appraised’; and *standard* as ‘a fixed reference point for use in assessing or describing the quality of something, or as a goal to aim for’. Sadler uses the example of testing a bicycle helmet: impact resistance would be a criterion; resistance at 75 units would be a standard.

- is fundamentally and transparently related to the set and taught curriculum
- has a suitable range and balance of subject content, learning experiences and assessment techniques
- has suitable task-specific criteria and standards for assessment
- has a suitable range and overall level of item difficulty and degree of discrimination
- is fair.

Additionally, teachers have the responsibility to expand, through design and use, the instructional utility of assessment instruments and reflect on and review critically with peers their assessment practices.

---

#### IV. Two key characteristics of externally moderated, school-based assessment

The system described in section III is not hypothetical. A working example of it exists in Queensland, Australia. In this section we explicate in greater detail two features of this operational system:

- the involvement of panels of teachers in commenting on and finally endorsing other schools' teaching and assessment programs (referred to as work programs) and reviewing standards in the assessment scripts of students in other schools
- student portfolio records.

Review panels are groups of experienced, practising teachers who give advice to schools in subjects within their area of expertise. Review panels in each subject area have a role in accrediting schools' work programs (helping schools develop courses that are consistent with the relevant syllabuses) and a role in reviewing submissions of student work (carried out in monitoring and verification meetings). Review panels operate at a local level (district review panels) and at a system-wide level (state review panels). State review panels are responsible for overseeing the work of the district review panels by sampling work programs and submissions across all local areas. The processes of *accreditation*, *monitoring*, and *verification* are fundamental in realising comparability of results between schools while allowing schools flexibility in implementing assessment. The processes are mandatory for all subjects in which statewide comparability of results is sought.

Accreditation is the process of periodic approval of schools' work programs. Among other things, work programs include an assessment plan that indicates the timing of assessment, the techniques to be used, the criteria to be assessed, and the ways in which performances in the criteria can be combined to decide levels of achievement.

Monitoring is the process, occurring towards the end of the first half of a course, of schools submitting the work of a sample of students to district review panels. Review panels provide schools with feedback on the quality of implementation of assessment and decisions about student achievement.

Verification takes place towards the end of a course. Schools submit the work of a sample (larger than the monitoring sample) of students, together with school decisions about students' levels of achievement. Review panels match standards in the student work with the criteria and standards of the relevant syllabus. If a review panel does not agree with a school's decisions, consultation and negotiations between the school and review panel take place<sup>11</sup>.

---

11. There are two other moderation processes: approval of schools' proposals for exit levels of achievement; and random sampling of student work. These are not discussed here as the former does not involve review panels and the latter is a *post-hoc* scrutiny of standards.

Moderation by review panels places bounds on the extent to which teachers exercise freedom in deciding standards of student work. In this way comparability of results is sought through consultation, negotiation and partnership between teachers and schools on the one hand and the broader system on the other. Consistent and valid moderation depends on standards, evidence and consensus; that is:

- the use of official statements describing standards at each level (which may be embedded in a criteria and standards matrix in which standards descriptors appear at each level for each criterion)
- the use of actual student work as the evidence upon which judgments are made about standards of student work
- the attainment of consensus among the judges that the work tendered matches the standards.

Fredericksen and Collins (1989, p. 30) suggest the need for a ‘training system for scoring tests’ in order to maximise reliability within an assessment system. A moderation process provides substantially more than mere training. Teachers not only submit their work programs and samples of the student work to review panels, they also consult and negotiate with review panels about the advice received from them. Where there is dispute, teachers engage in debate about matters such as the alignment of assessment programs with syllabus requirements, the appropriateness of assessment items, and the standards of student work. Information accumulated through the moderation process is fed back into the process of syllabus implementation and future development.

A student portfolio is a folder containing the student’s work in assessment tasks (scripts) set and graded by teachers in the school. The assessment tasks, implemented in accordance with the plan in the accredited work program, are a sample of the significant and mandatory aspects of the subject and include a range of assessment techniques, conditions, levels of difficulty, criteria, response types, and so on. Each student will have a portfolio for each subject. The portfolio is the evidence kept by the school in support of the school’s decisions about the student’s achievement in each subject.

Student portfolio records are, essentially, tables in which the rows are labelled ‘criteria’ and the columns are labelled ‘assessment tasks’. The standards reached (results obtained) by students in assessment tasks are recorded over time in cells of the table. The record may be annotated with descriptive statements of particular results. The record is usually on a single page attached to the front of the portfolio.

Two key aspects of the process of creating and maintaining individual student portfolio records are the ‘selective updating’ of assessment data and the use of the ‘fullest and latest’ information in determining levels of achievement (see Beasley 1987). These aspects set student portfolio records apart from data recording mechanisms that simply aggregate assessment results (and, of course, from terminal assessment, in which no account is taken of differences in performance over time). Selective updating refers to the process of reinterpreting assessment data in the light of new data generated later on in a course. The process is selective in two respects: not all new data provide a sufficient basis for altering previous assessment decisions and not all aspects or sections of a course are developmental in nature (and thus amenable to updating). Those data that relate to task-specific criteria that apply only to a particular aspect or section of a course are carried forward to the end of the course and then interpreted in terms of the global criteria and standards described in the syllabus.

At the time that students complete a course the profile contains a combination of data that have been carried forward in a relatively straightforward manner, and data that have undergone a process of updating and reinterpretation. The profile contains the fullest and latest information on the student's achievement.

Portfolio records are powerful tools of reporting. The power derives from the transparency of the profiling process and is a suitable example of the way *democracy* can be involved in valid educational assessment (see Cronbach 1988).

Each portfolio record empowers the student with knowledge of how the standard of a performance affects the overall result. As results accumulate, the student knows the standard reached in each criterion and the current overall achievement. The feedback provided is both diagnostic and remedial. It identifies areas of strength and weakness. The student can be involved in maintaining the portfolio record and can take action to improve achievement. In this way, school-based assessment encourages teachers to involve students in the decision-making process. Students know who makes decisions about achievement, when decisions are made, and on what information decisions are based.

---

## V. Contexts

A senior secondary school assessment system does not exist in a vacuum. It interacts with social, political, economic and cultural contexts. When we talk of assessment being 'high stakes', it is largely due to the influence of such contexts. There are many expectations arising from the various contexts. Amongst the many, some key expectations are:

- certification of student achievement
- articulation with the labour market
- articulation with tertiary education.

A key feature of a senior secondary school assessment system is how it accommodates contextual expectations. The choices made can be crucial. The 'wrong' accommodation can have severely distorting 'backwash' effects on the assessment system. As an example of how such contextual expectations can be met, we focus on the expectation of selection for tertiary entrance; that is, the allocation of (in some instances, scant) places in universities. This is a particularly important accommodation.

In the system we describe, teacher decisions about student achievement are recorded, without statistical adjustment, on a certificate issued by the curriculum and assessment authority. However, the allocation of university places, particularly in high-demand courses, requires finer distinctions between students. A statewide rank order of students according to overall achievement is constructed. The challenge is to meet this requirement without damaging the principles and values of externally moderated school-based assessment.

This is achieved by providing a profile of achievement information: a rank-order indicator of overall achievement; rank-order indicators of achievement in specific fields of study; results in subject-based school studies; a result in a cross-curriculum test. Profile-based selection proceeds in a step-wise manner that uses different information at each subsequent step. This approach is described in detail in Viviani (1990).

The abundance of information known to teachers implementing the system of moderated school-based assessment makes the construction of the profile relatively easy:

- teachers make fine-grained distinctions<sup>12</sup> between the achievements of students in each subject in each school
- fine-grained achievements are scaled<sup>13</sup> using the relevant group (*not* individual) results of the cross-curriculum test<sup>14</sup>, first to establish equivalence across subjects within each school, then to establish equivalence across schools.
- For each student, the processes of scaling provide a position of overall achievement (OP) reported as one of 25 bands, and positions of achievement in as many as five fields of study (FPs, reported as one of 10 bands)<sup>15</sup>. OPs and FPs can be used to make finer distinctions between students whose overall achievements in different combinations of subjects are about equal. Tertiary institutions can apply a step-wise selection approach using the profile of information as follows:
  - stage one—prerequisites (subjects and levels of achievement)
  - stage two—rank-order indicators of overall achievement (OPs)
  - stage three—rank-order indicators in relevant fields of study (FPs)
  - stage four—results in the cross-curriculum test
  - stage five—subsidiary information such as results in specific secondary subjects, testimonials, interviews or practical performances.

Every assessment system has to accommodate many requirements arising from the contexts in which it operates, some of lesser and others of greater importance than tertiary selection. The challenge is to accommodate the requirements without jeopardising the structures and procedures protecting the constructs, principles and values of the regime. In this particular example the solution, profile-based selection, enhanced the structures and procedures of the system by empowering teachers to make the fine-grained decisions needed for tertiary selection and confirming the ownership by teachers of the secondary school system.

---

## VI. Conclusion

We introduced this paper with a series of questions: ‘what do we believe in? what do we do? what is the connection between the two?’. We suspect that assessors specifically consider this series of questions (at least in the public realm) too seldom—a surmise borne out by comments in the literature. Madaus, Clarke & O’Leary (1997, p. 436) suggest that ‘testing is so ubiquitous that its role is taken for granted by most people’.

- 
12. Each student’s relative achievement (rank order and ‘gap’ compared to others) is decided.
  13. Scaling places the different sets of fine-grained achievements on a common scale. There are two stages of scaling: between groups of students in different subjects within a school (using subject-group results in the scaling test); and between schools (using school-group results).
  14. The test used for scaling is grounded in the curriculum. It is a test of the common elements of the curriculum, designed to be fully syllabus-based but not subject-specific, accessible to all students regardless of individual differences in subject choice. Unlike external exams, the test is driven by (rather than drives) the school curriculum (see Pitman 1993).
  15. While each subject contributes equally to the calculation of a student’s OP, subjects are weighted differently in the fields of study. Scaling for FPs involves the use of those parts of the scaling test that provide the best information in each field of study.

Johnston *et al.* (1995, cited in Black & Wiliam 1998, pp. 18–19) agree that the nature and role of testing is ill-defined, but take a far less sanguine view of the effects. Rather than taking assessment for granted,

... teachers ... were caught in conflicts among belief systems, and institutional structures, agendas and values. The point of friction among these conflicts was assessment, which was associated with very powerful feelings of ... insecurity, guilt, frustration and anger ... This ... suggests that assessment, as it occurs in schools, is far from a merely technical problem.

In working through the issues raised by our introductory questions we have identified and defined what we believe to be the key assessment constructs, validity and reliability, and have described the principles and values that inform, clarify and extend these constructs. These are that **high-quality assessment**:

- is integrated with learning
- is adequate, comprehensive and authentic
- produces fair and comparable results
- involves processes that allow for reflection
- is cost effective
- is accountable.

These constructs, principles and values are operationalised successfully in a system in which the structures and procedures involve assessment that:

- is grounded in good theory and practice
- is implemented and moderated in partnership between schools and a curriculum and assessment authority
- is school-based
- is subject-based
- values contextualised teacher judgments
- is continuous
- is based on criteria and standards
- is reviewed by panels of teachers
- makes use of student portfolios as evidence of achievement.

We propose that a high-quality assessment system will be able to accommodate contextual demands in a way that does not damage the links between its structures and procedures and its underlying constructs, principles and values.

The approach we have taken in this paper does not merely provide insights into particular assessment regimes and systems but is a basis for comparing regimes and systems. Parallel exercises undertaken in reference to other assessment regimes and systems would generate some interesting information about what unites and divides us as assessors.

Are we united on the key principles and divided on practice? Or do we split company on first principles? How well shared are our understandings of constructs such as validity? How important is it to situate assessment in the context of the learning environment? Do we value, devalue or distrust teacher judgments? An occasion such as the IAEA Annual Conference is an opportunity to explore these matters productively and so model hermeneutic practices.

---

## References

- Australian Curriculum, Assessment and Certification Authorities (ACACA) 1995, *Guidelines for Assessment Quality and Equity*.
- Beasley, W. 1987, *Profiling Student Achievement*, Queensland Board of Secondary School Studies, Brisbane.
- Black, P. and Wiliam, D. 1998, 'Assessment and Classroom Learning', *Assessment in Education*, vol. 5, no. 1, pp. 7–74.
- Broadfoot, P. 1992, *Assessment and Reporting at the Postcompulsory Level: Australian developments in the context of international trends*, IARTV Seminar Series, No. 13, April, Incorporated Association of Registered Teachers of Victoria, Jolimont, Victoria.
- Cronbach, L. 1988, 'Five Perspectives on Validity Argument', in Wainer H. (ed.) *Test Validity*, Earlbaum, Hillsdale, New Jersey, pp. 3–17.
- Elley, W. B. (undated) 'Examining school-based assessments: Some lessons from Canada', (unpublished paper?).
- Findlay, J. 1987, *Improving the Quality of Student Performance through Assessment*, Queensland Board of Secondary School Studies, Brisbane.
- Frederiksen, J. R. and Collins, A. 1989, 'A Systems Approach to Educational Testing', *Educational Researcher*, vol. 18, no. 9, pp. 27–32.
- Gipps, C. V. 1994, *Beyond Testing: Towards a theory of educational assessment*, The Falmer Press, London.
- Jessup, G. 1989, 'The Concept of Reliability in the Assessment of NVQs' in Jessup, G. 1991, *Outcomes: NVQs and the emerging model of education and training*, The Falmer Press, London, pp. 191–193.
- Madaus, G. F., Clarke, M. and O'Leary, M. 1997, 'The Kaleidoscope of Assessment: Disciplinary angles' (review essay: *Education, Assessment and Society* by P. M. Broadfoot), *Assessment in Education*, vol. 4, no. 3, pp. 431–438.
- Masters, G. N. & McBryde, B. 1994, *An Investigation of the Comparability of Teachers' Assessment of Student Folios*, Queensland Tertiary Entrance Procedures Authority, Brisbane.
- Matters, G. N., Pitman, J. and O'Brien, J. 1995, *An Exploration of Validity and Reliability that Tolerates Distinct Privileges: Standardisation and contextualised judgments*, paper based on two presentations at the twenty-first annual conference of the International Association for Educational Assessment, Montreal, Queensland Board of Senior Secondary School Studies, Brisbane.
- Matters, G., Pitman, J. and Gray, K. 1997, *Are Australian Boys Underachieving? An analysis using a validity–reliability framework based on the work of Lee Cronbach and Pamela Moss*, paper presented at the twenty-third annual conference of the International Association for Educational Assessment, Durban, Queensland Board of Senior Secondary School Studies, Brisbane.
- Moss, P. 1992, 'Shifting Conceptions of Validity in Educational Measurement: Implications for performance assessment', *Review of Educational Research*, vol. 62, no. 3, pp. 229–258.
- Moss, P. 1994, 'Can There be Validity without Reliability?', *Educational Researcher*, vol. 23, no. 2, pp. 5–12.



O'Brien, J. 1999, 'The Framework in which Assessment of Senior Secondary Mathematics is Implemented in Queensland', in Baldwin, K. and Roberts, J. (eds) *Mathematics: The next millennium*, papers from the Seventeenth Biennial Conference of the Australian Association of Mathematics Teachers, Inc., pp. 42–46.

Pitman, J. A. 1993, *The Queensland Core Skills Test: In profile and profiles*, paper presented at the nineteenth annual conference of the International Association for Educational Assessment, Mauritius, Queensland Board of Senior Secondary School Studies, Brisbane.

Pitman, J. A. and Dudley, R. P. 1985, *Criteria-based Assessment: The Queensland experience*, paper presented at the eleventh annual conference of the International Association for Educational Assessment, Oxford, Queensland Board of Senior Secondary School Studies, Brisbane.

Sadler, D. R. 1986a, *ROSBA's Family Connections*, Queensland Board of Secondary School Studies, Brisbane.

Sadler, D.R. 1986b, *Subjectivity, Objectivity and Teachers' Qualitative Judgments*, Queensland Board of Secondary School Studies, Brisbane.

Viviani, N. 1990, *The Review of Tertiary Entrance in Queensland*, report submitted to the Minister for Education, Queensland Board of Senior Secondary School Studies, Brisbane.