
A Submission to the Parliament of Queensland's Education and Innovation Committee: The Assessment of Mathematics, Physics and Chemistry

Submitted by: Peter Philip Thomas
Address: [REDACTED]
Position: Former Head of Department, Modern Languages,
Toowoomba State High School
Formal Qualifications: B.A., B.Ed. (U.Q.); Litt. M., M. Ed. (U.N.E.); M.A.C.E.
Teaching Experience: 33 years in secondary education (now retired)

SMC&PA Submission 39
Received: 29 April 2013

Term of reference: **The ability of assessment processes to support valid and reliable judgements of student outcomes**

Preamble:

This short submission to the Parliamentary Education and Innovation Committee's inquiry into the assessment of mathematics, physics and chemistry will, I hope, **prove useful mainly to those committee members whose interests and backgrounds in education are of a non-professional nature**. I am sure that the Committee will receive submissions containing the detailed technicalities of both sides of the current dilemma from their proponents. By contrast, this expresses in uncomplicated terms the background to the debate, and examines some of the implications now arising from it. It is primarily for the "lay" reader as a prelude to the more complex considerations which must follow.

It should be noted that it is presented by a teacher and administrator who is now retired, but who has had many years of experience in teaching and assessing secondary "humanities" subjects, including foreign languages, English, and geography.

Background:

Historically, the assessment of all school subjects, primary and secondary, was carried out by the allocation of marks, usually given out of 10 or 100. Committee members whose school-days were before the mid-1970's will well remember the mark at the bottom of their work, circled in red ink, with a succinct comment alongside. In longer examinations, these marks were usually added to give a percentage, and depending on the band where one's mark fell on this 100-point scale, a letter grading – A, B, C, etc. was sometimes added. In the Junior and Senior external examinations, only the letter grading was provided to students.

The apparent advantages of the "percentage" marking system were manifest. If the test was a "right or wrong" one, e.g. spelling or mental arithmetic, the teacher was easily able to calculate a mark, and this could be readily understood by students, parents, other teachers, and school inspectors. Not only did it provide information on individual abilities, it was able to give an indication of the student's performance with respect to the rest of the group. If marks for the whole group were significantly depressed or elevated, it could provide a comment on the standard of the test - or perhaps of the teaching!

If the test was a written composition, or perhaps the painting of a picture or the recitation of a poem, the same marking scheme was historically applied quite confidently by most teachers. However, whether they knew it or not, a new dimension was now present – subjective judgement. In other words, the mark allocated to the work depended to some degree upon the teacher's personal opinion of it. The simple "right or wrong" test no longer applied. Certainly, a word could be mis-spelt in a composition or a line forgotten in a poem and such errors could be objectively tallied up, but such mechanical aspects made up only a part, and perhaps quite a small one, of what was being tested. The rest depended on personal taste and opinion, and a second teacher might give quite a different assessment of it, as might a third. Thus, the validity and reliability of such marking could be called into question.

The allocation of marks to such "value-judgement" student work reached a ludicrous level at eisteddfod performances, where it may still be found alive and well. I am sure that many committee members have attended such performances and sympathised with the distraught eight-year old who has played a piano solo beautifully for 87 ½ marks, only to be beaten into second place by a half-mark. What mortal adjudicator could allocate marks for an artistic performance on what is essentially a 200 point scale, and having performed that miracle, then separate two competitors by a single mark?

Transition:

By the 1960's, most researchers were agreed that subjective assessment of "value-judgement" work might be done on a five-point scale at the very best. Research demonstrated that markers, no matter how experienced, were not able to discriminate such work into more than five broad bands, thus making a nonsense of allocating a percentage mark (i.e. division into 100 bands) to it.

The implications for the "humanities" subjects were clear. The 1970's began a long transition from marks and percentages to evaluation based on the student's satisfying, to a greater or lesser extent, a number of agreed *criteria*. Five broad bands, ranging from Very High to Very Low, were proposed. Matrices of descriptors were prepared to show the student's level of achievement in satisfying the criteria in tests which were set to match them closely. Examples of these matrices are not included in this submission - they may be examined in the published subject syllabuses.

There were advantages in criteria-based assessment over marks in the "humanities" subjects. Because their evaluation was largely subjective, the fallacy of awarding marks and percentages was avoided, and, provided the wording of the descriptors was lucid and the bands clearly defined, such assessment provided a useful normative and diagnostic tool for both student and teacher. Although an older generation of parents probably preferred the percentages of their own school days, most came to accept the new system, as the descriptors gave a detailed commentary on their child's work and progress.

On the other hand, criteria-based assessment presented difficulties. Firstly, if the descriptors were not carefully and simply written, ambiguities and other problems in comprehension arose. I recently perused the descriptors for Year 3 Mathematics. I regard myself as reasonably literate, and have been exposed to "educalian" for over 30 years, but in spite of that found myself at a loss to understand, in many places, exactly what was meant. I believe that a first-year teacher would have rather more difficulty than I did. In addition, even if the descriptors are fully understood, many teachers would find it difficult to interpret and apply them accurately to student work. Put simply, some of the original problems posed by percentage marking have resurfaced in another guise.

In spite of these problems, my experience of both systems leads me to agree that criteria-based assessment is a suitable and useful tool in the evaluation of most of the content of the "humanities" subjects.

Of babies and bathwater:

In its eagerness to embrace criteria-based evaluation, the assessment authority made, in my view, a serious error in declaring that the use of criteria and descriptors should be the major tool in the assessment of **all** school subjects, not just those where value-judgements were required. Although there is a certain tidy uniformity in having assessment matrices looking the same across the entire range of subjects offered, that is about all that can be said for such a standardised programme.

It is perfectly possible to write assessment instruments in mathematics which can be marked objectively. Indeed, the vast bulk of primary and secondary maths readily lends itself to this form of marking, whether one is evaluating processes used by students in problem-solving or the knowledge of algorithms, theorems and number facts. Part-marks may be confidently allotted where some, but not all of a process has been carried out, as it is usually quite clear-cut where the student's reasoning has broken down. It is therefore possible to give a numerical or percentage mark which is a very accurate assessment of a student's handling of a particular task. There is simply no real need for a matrix of descriptors because, at least for mathematics and most science, it cannot and does not provide the precision of evaluation offered by numerical marks.

As a humanities teacher and administrator, I believe the move to criteria-based marking in the humanities was necessary and long overdue. However, there was simply no need to apply the one-size-fits-all approach to those subjects where it was not only unnecessary, but indeed counter-productive.

Proposed: a dual system:

Although it runs counter to the assessing authority's fetish for strict uniformity in the method of student evaluation across the entire range of subjects, I believe that there are several advantages in permitting a dual system of assessment. Those subjects where objective marking is not possible should be required to use the criteria/descriptor system currently mandated. There can be no doubt of its suitability and superiority here. On the other hand, those subjects such as mathematics which can be readily assessed by marks and/or percentages **should be given that alternative**. The placement of students in bands of achievement resulting from such marks could be readily agreed upon, I believe, without returning to external examinations. In any case, it would present no greater difficulties than those currently experienced in implementing the criteria-based method.

One might even argue for such a dual system to be used *within each subject* where appropriate testing instruments of both kinds were used, but at this point, such a heresy is quite insupportable because of the difficulties in implementation and the potential scope for unfairness to students in different schools.

Conclusion:

Although the criteria/descriptor form of evaluation has a useful and necessary place in the assessment of work which requires multiple value-judgements on the part of markers, the mark/percentage method is manifestly more suitable for such subjects as mathematics and the sciences. It is more easily interpreted by markers and consequently easier to apply to student work. It is more precise in its message, and therefore more readily understood by students and parents. Where necessary, it might be supplemented by a descriptor commentary on the student's use of various skills, but this would be informative, rather than evaluative.

Each method has its advantages, and in my view, there is no good reason why the two cannot co-exist under the assessment umbrella, each being applied where it is best suited.

27th April, 2013