SMC&PA Submission 69
Received: 7 May 2013

Re: Assessment Methods for Senior Maths, Chemistry and Physics

I am making this submission at the prompting of a friend who has been a mathematics teacher in Queensland for many years. Although it is something I would rather not do I should also perhaps say where I am coming from. I have both taught mathematics, and researched its teaching and learning at high school and university levels, and been a plenary speaker at international conferences sponsored both by education groups and by mathematicians. Assessment wise I have been a marker within the external examination system of another state, and the inaugural Chairman of the A3 Mathematics State Review Panel which oversaw the introduction of the ROSBA system in Queensland. I have served on the Executive Committee of the International Commission on Mathematics Instruction (ICMI) and as President of the International Conference for the Teaching of Mathematical Modelling and Applications (ICTMA). Among other roles I was an editor of the Study volume that contained the proceedings of the 14th ICMI Study on Modelling and Applications in Mathematics Education, and an invited contributor to an earlier ICMI Study on investigations into assessment in mathematics education.

While not presently involved in field activity I have tried to keep up with the current debate concerning mathematics assessment in Queensland, through interaction with friends and colleagues in the field and attention to such media reports as have surfaced from time to time. From this background I offer the following comment.

Valid school assessment needs to follow the same principles that apply to measuring excellence in the world at large. As a non-academic example take the Olympic Games. In some events (for example 100metres and marathon) it makes sense to put all competitors on the track at the same time, with the first to cross the finish line being the winner and so on down the order. One simple criterion, one simple measure (mark)!

It makes no sense at all to apply such a method to events like diving and gymnastics, which involve individual choices with different degrees of difficulty. Here there are multiple criteria in the heads of a panel of judges, and there is some degree of subjectivity in the assessments, as each assigns a ‘score’ to performances. Interestingly (political influences aside), there usually seems to be a pretty good consensus regarding outcomes.

Closer to home the same principles apply within the discipline of mathematics itself. Undergraduate subjects that involve mastery of basic content, skills, and applications, are commonly tested through timed examinations, perhaps with augmentation from some other sources such as tutorial work.

But when it comes to creative work such as theses (and journal articles) a totally different approach is used. No one would suggest that in order to properly and fairly judge and compare the quality of mathematical theses every candidate in a particular field should be given the same topic, and the same date for submission – it would be theoretically indefensible as well as impossible in practice. In the case of theses and other synthetic work a sense of what is appropriate is inbuilt into the culture of the discipline and its practitioners. Appropriate criteria are carried in the heads of

examiners, as is the sense of the standards to which the criteria are met in different cases. If supervisors have done their job properly different theses will all contain correct mathematics - but still be judged as differing in quality – without a mark in sight. The issue has nothing to do with the level of the mathematics as such, but rather the nature of the kind of activity being considered.

Returning to school mathematics the job of assessment should not be the onerous one that is suggested by some of the comments and examples that have been cited, involving myriads of checked boxes, and consuming unnecessary time and effort.

I see the issue as maintaining validity and efficacy but eliminating unnecessarily cumbersome assessment procedures - that eat into time and energy for no additional benefit. This means, for example, using testing and marks for those aspects of mathematics that can be validly tested by these means. It seems to me that those protesting the use of marks may not recognise that qualitative criteria with awards such as A, B, C etc have implied underlying ordinal grading scales associated with them.

Test questions can be used to assess much more than the routine kinds of mathematical knowledge that some opponents like to pretend. For example: resolve conflict; compare/contrast; justify; create; identify and correct error; defend; spot an anomaly; look before you leap.

On the other hand the proposed Australian Curriculum in mathematics indicated in its framing statement that "mathematics aims to ensure that students are confident, creative users and communicators of mathematics, able to investigate, represent and interpret situations in their personal and work lives and as active citizens". It now goes on to say that: "It (the national mathematics curriculum) develops the numeracy capabilities that all students need in their personal, work and civic life, and provides the fundamentals on which mathematical specialties and professional applications of mathematics are built... These capabilities enable students to respond to familiar and unfamiliar situations by employing mathematical strategies to make informed decisions and solve problems efficiently.

Australian Curriculum: Mathematics Rationale. Retrieved March 16, 2013 from: http://www.australiancurriculum.edu.au/Mathematics/Rationale

To seriously address capabilities like these (that include abilities such as problem solving, applications, and modelling) formal testing is inappropriate – indeed invalid, an understanding that has widespread international recognition, including within leading groups such as ICMI and ICTMA. However this does not justify the setting of time consuming assignments that seem to work on the principle that if someone writes enough they are more likely to cover what is required. (And even more relating to reports, if true, that in some assessments, students are diverted to writing about the role of the subject in society, rather than engaging with its content.) As far as assessing students' abilities to apply mathematics in application contexts (real not contrived) there are various approaches that value parsimony as well as thoroughness. One such approach that has been used for modelling tasks involving model formulation, making assumptions, solution of mathematics, interpretation, evaluation and/or refinement involves the construction of a poster as a report – with size limited to two sheets of poster paper. Sometimes this has been supplemented by a formal succinct supervised written component where the student then answers questions directly related to their project. (This is a kind of parallel to the oral defence of a

thesis which is another check that the individual student actually understands what they have written - and indeed have done the work). A variety of assessment approaches have been documented as they have been used in a variety of countries.

So my underlying position is to reiterate the rather obvious imperative that assessment procedures should be constructed to validly and efficiently measure the particular type of knowledge that is the object of interest, and do so fairly and reliably. This is the prime consideration, rather than the popularity of particular approaches in other places – reasons for which will vary. Common testing provides an efficient means of assessing certain types of knowledge, but other important forms of knowledge cannot be validly tested in this way. I am concerned that in debates of this nature positions can become entrenched, and it is always possible to provide horror stories of happenings within any form of assessment that one does not favour – I could certainly do so for both examinations and assignments.

In summary inadequate assessment methods should not be defended, whether they involve cumbersome and irrelevant assignment work, or examinations claiming more than they are capable of measuring. We need to focus on improving the quality and efficiency of instrumentation to ensure that teachers' time is not diverted and their energy and patience further eroded. I hope that this present debate does not descend to the simplistic level of setting one genre of assessment against another. In particular, if concern about mediocrity in some aspects of the system is a concern, then imposing any type of 'one size fits all' approach would simply replace one mediocrity with a different one.

Honorary Professor Peter Galbraith

[redacted]