

# Authentic assessment: Its application to mathematics education

---

Elizabeth McCabe

*“Authentic assessment has a lot to do with acknowledging the personal, social, cultural and political contexts of assessment and relating them to what happens to people in assessment situations.”*

In this article I look at the subject of authentic assessment and its issues. The latter part of the paper uses examples drawn from personal involvement in mathematics education in two sectors, primary and tertiary. I shall draw out and discuss selected aspects of authentic assessment.

## What is authentic assessment?

Many mathematics educators now believe that the possession of isolated skills can no longer be equated with an ability to ‘do mathematics’ or ‘be a mathematician’. Evidence of rich and integrated skills and knowledge is now desirable. Accordingly there is a push for assessment philosophy and practice to change. Also, the search is on for a fairer picture of student understanding (Davis, 1995; Clarke, 1995). Clarke (1993) describes this emphasis and some associated actions on its behalf, as a call for ‘authentic [over traditional] assessment’. An important part of placing ones hat in the authentic assessment ring is recognising the imperfect nature of all assessment - ‘authentic’ or otherwise. The fact that a particular context, task, technique or pedagogy (e.g. journal writing) is sometimes promoted alongside discussions of authentic assessment can be misleading in this respect. Such promotion should be viewed as an attempt to take assessment beyond thin procedures which check isolated skills rather than as a means to model perfect pedagogy or assessment practice (Davis, 1995; Clarke, 1992). Searches for exemplary assessment methods, tasks or

approaches do not go with authentic assessment.

A key objective of authentic assessment is the minimalization of the disadvantage(s) in any given assessment situation or assessment context (Clarke, 1993). There is plenty of evidence to suggest that such disadvantages exist. Much literature on the subject of poor performance in mathematics highlights the cultural, political and social differences in viewpoint that may exist between assessment designers (e.g. classroom teachers, state educational bodies, test publishers) and those assessed (see for example Bishop, 1992). Authentic assessment has a lot to do with acknowledging the personal, social, cultural and political contexts of assessment and relating them to what happens to people in assessment situations. The push toward authentic assessment, and assessment in similar vein (e.g., Realistic Mathematics Education assessment techniques from the Netherlands as characterised by van den Heuvel-Panhuizen, 1996) has involved challenging the credibility of, and traditional balances between, the psychometric quality controls of reliability and validity. It has also involved seeking acceptance of qualitative over quantitative data. Thus would-be authentic educators swim against a tide of past assessment standards. My experience has been that it takes dogmatic professionalism, over an extended period of time, to do so.

To my mind authentic assessment ideals must be tempered with a touch of reality. Unless the endpoint uses for assessment(s) ing



*Elizabeth McCabe teaches at the Auckland College of Education campus in Whangarei. Although she specialises in mathematics education, she also teaches a number of other curriculum courses.*

change, unless assessment by portrayal rather than by measurement is accepted by many, unless many educators, students and members of the public acknowledge assessment as much for what it is not as what they believe it to be, progress in overtaking traditional assessment genre will be slow.

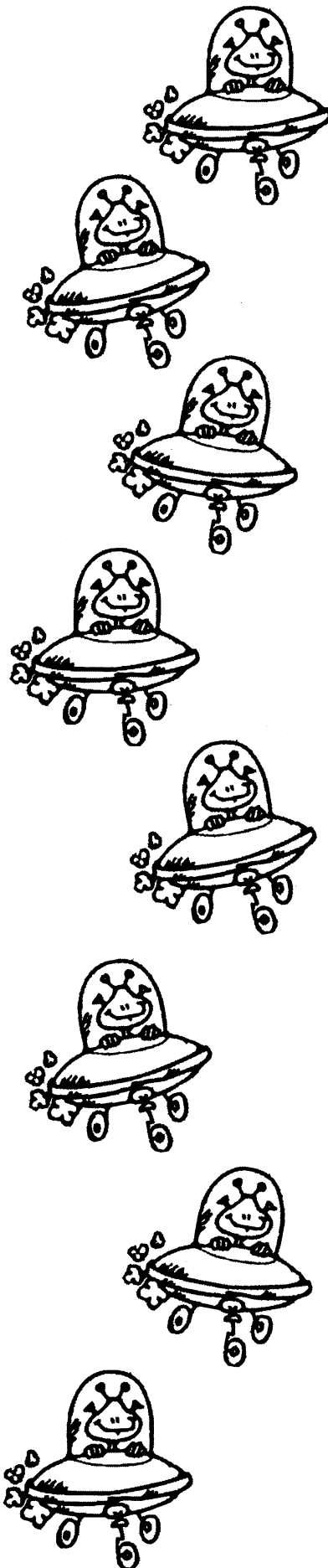
### Authenticity

Authentic assessment situations will empower teachers, and caregivers as well as pupils. Assessment information is perhaps only authentic when the classroom teacher possesses both the confidence and ability to use that information on the students' mathematical behalf (Clarke, 1995; NCTM, 1995). A popular [constructivist] view is that a teacher is more able to do this when he/she is party to rich and sometimes complex kinds of assessment data in which a student's perspective of the task, cultural and otherwise, is both acknowledged and accounted for in a genuine way (Clarke, 1992; Arlø & Skovsmose, 1996)

A deaf child may do better on a traditional pencil and paper test than in an interview. What seems to be important in authentic assessment is to take into account are the many factors which contribute to the final assessment situation. A student's prior knowledge, his/her familiarity with the testing situation, the contexts real or otherwise within the assessment task, the cultural background of the teacher, and the cultural background of the student's family may all impinge on an assessment situation.

### The contextualisation of mathematical tasks

Imagine giving a problem about aliens to a class of children. You tell them the aliens want to calculate their fuel requirements and therefore stopping places for a journey. The sticking point for students may not be that they are fantasy creatures. (From my classroom experience most children are happy to enter into a fantasy world replete with imaginary beings) It may be



that a number of children in the class have little familiarity with the everyday context of refuelling the car. The use of an appropriate context is felt to be important in making authentic assessment (van den Heuvel-Panhuizen, 1996). Using a broad definition, an appropriate context can be described a real-life [my preference], imaginary or mathematical one which allows a student to personally connect with elements in a given assessment task (Boaler, 1993; van den Heuvel-Panhuizen, 1996).

It is my opinion that student involvement in assessment is a much neglected assessment direction. In my class children were assessed on their ability to measure accurately while they built their own kaleidoscopes. Self assessment came naturally to a high percentage of my class as they looked to improve their product. I would describe this as the rightful [and authentic] purpose for assessment (Clarke, 1992 and Stenmark, 1991 express similar views). There are other ways students can [and I believe should] become involved in assessment. As yet student involvement in assessment design is not a common practice and I believe it should be.

*“...student involvement in assessment is a much neglected assessment direction.”*

### The empowerment of the learners - cultural considerations

A number of writers and researchers are of the view that unless the potential dissonance between the cultural values and background of learners and their assessors [or the teaching or assessment contexts, tasks and methods tasks or philosophies they subscribe to] is recognised, teachers will continue to come to unfair conclusions about both the ability and academic potential of

*...cont'd on page 47*

their students (Bishop, 1994).

The advent of Pangu i roto i te marautanga o Aotearoa (Ministry of Education, 1997) a mathematics curriculum by Maori for Maori should keep authentic assessment issues to the fore in New Zealand. The creation of this curriculum was predicated on a desire for validity in both the teaching and assessment of Maori students (Ohia, 1993). The historical relationship [between the participants in the assessment process] in the case of Maori has been that the medium of both instruction and assessment has been the English Language since the early 1900's. When representatives of Maoridom came to take stock of the educational needs of Maori people the one-sidedness of the origin and ownership of the mathematics being taught and assessed in classrooms was challenged. Power relationships were discussed (ibid.). Authentic assessment looks at exposing or ameliorating perceived imbalances in an assessment process, cultural or otherwise. The question for the authentic assessor is what imbalances can or should be tolerated. This is a question that a group of Maori educators and other interested parties have started to debate. I believe one of the challenges for the would-be-authentic assessor in the classroom is to go beyond the use of teaching and assessment contexts with cultural veneers, for example enlarging a koru (Boaler, 1993). The challenge is to consider some of the deeper issues in assessment such as power relationships and the politics which sustain them.

**The empowerment of the learners , opportunities to express the outcomes, the conditions in which an assessment takes place**

I remember introducing a problem solving assignment to a class of pre-service teachers. My students and I sought "a consensus conception of competence" (Clarke, 1992) with some difficulty. Open-ended questions seem to afford



opportunities for students to display the richness of their mathematical understandings. However there may be problems with the format of open-ended questions because they do not indicate clearly to students the cognitive level, or desired content of the response required (van den Heuvel-Panhuizen, 1996). Somehow then any incongruences between the underlying intentions (perhaps to test content knowledge in a particular domain) of the tester and the freedom given to student to select their own investigative direction must be dealt with. Scoring rich-result or authentic-type assessment can also cause trouble. Descriptors used to grade levels on performance tasks have been found to be problematic. Swan's (1996) British attainment target assessment experience has been that "Many tasks do not permit students to display performance at a full range of levels" and "If students are unaware of which aspects of performance are being assessed they are unlikely to display these aspects." Multiple assessment tasks, designed to elicit a gamut of mathematical behaviours, are one alternative. Clarke (1993) describes a multiple-component assessment system used to assess the mathematical ability of 12th graders in Victoria, Australia. Davis (1995) coins the term 'a basket of different kinds of evidence' to describe the conglomeration of assessment clues e.g. observations, tests, interviews and project work now accepted by some as legitimate evidence of understanding. Nowadays students may have the option of performing in a range of conditions. Truly a headache for one with a simplistic assessment perspective!

In 1993 in my Year Three and Year Four Classroom the multiple components of my assessment system in mathematics included child self-review through mathematical diary writing, the keeping of annotated class lists, anecdotal report writing on a formal school report, portfolio samples of both individual and group work, one formal test, a few teacher made tests and response sheets ( includ-

*“To my mind it can be accorded at least the status of a test score provided the context in which it occurred is described”*

one or two asking children to appraise a peer).

Here is an entry from 7 year old A's diary.

*Today we did water work. I learnt that five cups go in a litre bottle and I thought it was two and a quarter cups. Mrs A [teacher aide] liked my work. She said it was excellent. The difference was two and three quarters. It was quite easy work.*

A demonstrated abilities and understandings that reached well beyond the current level of instruction. Information that a traditional assessment measure would probably not have picked up. I affirm Clarke's (1992) belief that informal or semi-formal information like this has value for both instruction and assessment. To my mind it can be accorded at least the status of a test score provided the context in which it occurred is described (e.g. her mathematical method cannot be implied from her writing). At that time I would have had no hesitation in using it as professional evidence in a formal setting such as a parent interview or a visit from the Educational Review Office.

### **Validity, Reliability, Endpoint Uses of Assessment, Measurement and Portrayal**

There is an idea that I like. It was mooted by Freudenthal (van Heuvel-Panhuizen, 1996). It is the idea of “well-designed intersubjectivity” and Freudenthal proposed it “supersede ill-conceived objectivity”. Van den Heuvel-Panhuizen (1996) says this:

*In contrast to the past, when teachers were expected to disregard any knowledge they*

*held of students while correcting their work, they are now increasingly urged to use this knowledge.*

She describes a growing recognition among some members of the mathematical education community that the pursuit of validity of assessment, more often than not diminishes the measurable reliability of that same assessment. She also claims that in the current educational climate validity should frequently be viewed more desirable and more meaningful than reliability. She suggests an incompatibility exists between the construct of reliability, and the educationally embedded purposes of assessment. Like Davis (1995), van den Heuvel-Panhuizen notes that the complexity and originality of the mathematical responses sought today, and the methods by which such responses can be obtained and evaluated do not fit in well with a search for reliability. The problem seems to be mismatch between the philosophy underpinning present directions in the assessment and teaching of mathematics, and the philosophy underpinning the use of reliability (itself a technicist construct) as a measure of the success of an assessment procedure. Davis suggests, and I agree, that efforts to squash new-style assessment procedures to fit the mould of old ideas of comparability, embodied in reliability scores and ranking procedures, are futile.

*“...efforts to squash new-style assessment procedures to fit the mould of old ideas of comparability, embodied in reliability scores and ranking procedures, are futile.”*

### **Consequent action**

I recall the range of responses by parents as my school moved from a grading system for most subjects,

*“Many parents liked the new system but it was evident that some parents still sought the comparability and comfort of formal, but often by nature less in-depth, scores”*

mathematics included, to a report which was a collection of dated comments. An adjunct to the report was a portfolio of the child's work often with evaluative remarks by the child, a peer or the teacher.

A dated report comment might look something like this:

*6/7 Can mentally work out the distance between ball throws (e.g. 43 m and 36 m) and also record the throws as subtraction equations.*

For the teachers the recording was time consuming but it had advantages. The frequent frustration of writing the same evaluative comments in many places for different audiences was diminished. The contextual nature of the comments built up an interesting picture of the situations in which a child could perform certain type of procedures. Some teachers began annotating portfolio sheets with comments like “with help” or “with a prompt”. They found this useful when defining progress, and deciding future directions.

The ability to inform consequent action is a key element of authentic assessment. The teachers above could use assessment information for both planning and reporting. Well attended interviews and explanatory meetings helped caregivers understand what was going on and why, and also how their child was doing. Many parents liked the new system but it was evident that some parents still sought the comparability and comfort of formal, but often by nature less in-depth, scores from procedures like international, national or school based tests.

Stenmark (1991) mentions how important it is for teachers to bring parents and the community 'on-side' with new ways of looking at assessment. Like Stenmark I feel the ability to keep up the lines of communication between all participants is crucial to the success of such endeavours.

### Endpoint Assessment

In my view endpoint assessment is most problematic for authentic assessors. This is because the authentic assessor may be asked for a broad and traditional kind of generalisation based on a set of idiosyncratic information. I for one am reluctant to see my rich information distilled into a grade, something I seldom faced when teaching in a primary school, but face every day teaching in a tertiary institution.

*"I hope that advocates of authentic assessment will neither procreate nor feed a fiction that assessment can be fair to all."*

### Learning Theories and Paradigms

Empowerment, and possibly also constructivist paradigms fit poorly with the idea that an expert or experts, even with some degree of agreement, can know or assess the quality of someone else's understandings. More Vygotskian (Lunt, 1993), or perhaps strongly socio-constructivist principles (Ernest, 1994) would see teachers ready to accept the mantle of professional expert with a concurrent responsibility to interpret student understandings and act accordingly. To my mind, in spite of it's inherent difficulties to do with subjectivity (Swan, 1996), [for example the multiple constructions by teachers of any given task or the reasons for setting it (Mousley, Clements & Ellerton, 1992)], developing and improving criterion-based consensus moderation is a better path than returning to the

testing of isolated skills in formal examinations or tests. As Davis (1995) suggests although assessing rich knowledge may be troublesome abandoning the goal is unpalatable.

### Conclusion : Answering the call for authentic assessment

I hope that advocates of authentic assessment will neither procreate nor feed a fiction that assessment can be fair to all. What stands in the way is the philosophical impossibility, especially if you are a constructivist, of knowing another completely. The reality is that even tailor-made assessment for one individual cannot be truly authentic. Even so, current and future searches for 'better problems' may give us mathematical assessment tasks that are personally, culturally and academically fairer than what we have had before. Problems in which the prior and present experiences and understandings and cultural backgrounds of the students are taken into account. Problems which, in a best case scenario, avert teacher misinterpretation of pupil responses. [Safety net problems developed in the Netherlands, and detailed by van den Heuvel-Panhuizen (1996), are but one example of an attempt to overcome this difficulty. A first question is open, while a second question focuses the response into a more clearly defined domain].

The underlying idea of authentic assessment is that educators should not be smug about how and why they assess. The issue of ownership of, and responsibility for assessment be it by a cultural group (e.g. New Zealand Maori), an individual (e.g. self-assessment) or a community of practice (e.g. teachers and mathematicians) overarches the authentic assessment debate. I think authentic assessment describes the appearance, of a timely set of assessment principles and derivative actions. For me the most important thing about Clarke's (1993) call for authentic assessment, is that no amount of answering can make it go away.

### References

- Arlø, S., Skovsmose, O. (1996). Students' good reasons. For the Learning of Mathematics ,16,3, 31-38.
- Boaler, J. (1993). The role of contexts in the mathematics classroom : do they make mathematics more "real"? For the Learning of Mathematics , 13,2,12-17.
- Clarke, D. (1992). The role of assessment in determining mathematics performance. In Leder, G. (ed.) Assessment and learning of mathematics . Hawthorn, Australia: The Australian Council for Educational Research.
- Clarke, D. (1993). The language of assessment. In M. Stevens et al. (Eds.) Communicating Mathematics .
- Clarke, D. (1995). Constructive Assessment: Mathematics and the student. In A. Richards (Ed.) FLAIR - Forging Links & Integrating Resources . Proceedings of the 15th Biennial Conference of the Australian Association of Mathematics Teachers. Darwin: Australian Association of Mathematics Teachers.
- Davis, A. (1995). Criterion-referenced assessment and the development of knowledge and understanding. Journal of Philosophy of Education, 29,1, 3-21.
- Ernest, P. (1994). The philosophy of mathematics and the didactics of mathematics. In R. Biehler; R. Scholz; R. Sträßer, and B. Winkelmann (Eds.) Didactics of mathematics as a scientific discipline . Dordrecht: Kluwer Academic Publishers,
- Lunt, I. (1993). The practice of assessment. In H. Daniels (Ed.) Charting the agenda, London: Routledge.
- McCabe, E. (1998). 0781.533S journal (unpublished).
- Ministry of Education (1992). Mathematics in the New Zealand curriculum . Wellington, NZ: Learning Media.
- Mousley, J., Clements, M., Ellerton, N. (1992). Teachers' interpretations of their roles in mathematics classrooms. In Leder, G. (ed.) Assessment and learning of mathematics . Hawthorn, Australia: The Australian Council for Educational Research.
- NCTM (1995). Major shifts in assessment practice. Assessment standards for school mathematics . Reston, Virginia: National Council of Teachers of Mathematics
- Stenmark, J. (1991) (Ed.). Mathematical assessment - Myths, models, good questions and practical suggestions . Reston, VA: NCTM.
- Swan, M (1996). Assessing mathematical processes: The English experience. Mathematics Teaching in the Middle School ,1,9,707-711
- van den Heuvel-Panhuizen, M. (1996). Assessment and realistic mathematics education . Utrecht: CD-8 Press, Center for Science and Mathematics Education.

