

Criteria for authentic assessment of mathematics: Understanding success, failure and inequality

Candia Morgan

Institute of Education, University of London, UK

The notion of the desirability of authentic assessment has become widespread across many areas of the curriculum and at various levels of education, including higher education (Torrance, 1995). Within mathematics education, there has been considerable international interest in discussing the characteristics of more authentic means of assessment and attempting to design and implement them (see, for example, Niss, 1993a, 1993b; Romberg, 1995). The term *authentic*, when used in the context of curriculum and assessment, has rich and multiple connotations that are not always distinguished from one another or clearly analysed. Cumming & Maxwell (1999) provide a useful categorisation, identifying and distinguishing four major interpretations of authenticity, each based on different underlying theories of learning: performance and performance assessment; situated learning and situated assessment; complexity of expertise and problem-based assessment; competence and competence-based assessment. Within the mathematics education literature, however, the underlying theory of learning is usually left implicit and the term *authentic* is used as if it were unambiguous. In this paper I intend to explore the meaning of authenticity in relation to recent changes in mathematics curriculum and assessment. In particular, I shall consider the forms that assessment criteria may take and the implications for both individuals and groups of students. The examples I employ relate specifically to the use of 'open-ended tasks', introduced in many curriculum reform movements internationally (Pehkonen, 1997). The principles for interrogating the nature of assessment criteria and their implications are, I believe, more generally applicable.

At the level of the assessment task itself, authenticity tends to refer to the nature of the mathematical activity involved — but this may be judged according to some idealised version of the nature of mathematics (for example, 'Mathematics is a creative activity, so authentic tasks should allow space for students to be creative rather than to repeat memorised procedures'), relative to a view of what students will find meaningful, often involving some 'realistic' aspect or, according to Cumming & Maxwell (1999), reflecting adult activity. Authenticity is also often considered at the level of the entire assessment system. In this case, it is usually taken to mean that assessment procedures match the aims, content and breadth of the curriculum; the assessment system is thus an authentic representation of the curriculum it is assessing. This aspect of authenticity is often asso-

ciated with what is called ‘performance assessment’, that is, methods of assessment that are based on assessment of students’ performance on tasks that they undertake as part of their everyday classroom learning activity (portfolio assessment, in which students form a collection of examples of work done during a course in order to display the range of their achievement, is a good example of this) or tasks that, although designed specifically for assessment purposes, mirror the kinds of tasks that are considered to be part of ‘good practice’ in teaching. In the context of curriculum reform movements where assessments are designed externally rather than by teachers themselves, this matching of assessment to desirable curriculum and pedagogy has been seen as a powerful tool for furthering reform, recognising that teachers are likely to match their teaching to what they know is to be tested (Bell, Burkhardt, & Swan, 1992).

The rhetoric of authenticity assumes that the results of such assessment are a more genuine measure of students’ mathematical achievement than those traditional forms of assessment to which they are being opposed. But who is arguing that assessment should *not* be authentic? Certainly many proponents of traditional forms of assessment in mathematics in the United Kingdom, though not using the term *authentic* themselves, argue that the kinds of tasks they prefer to use (mainly closed questions demanding recall and application of standard procedures, to be answered by students working alone within restricted time limits) test the aspects of mathematics that they value (for example, acquisition and accurate deployment of standard methods and notations) and that the time pressure that students are put under in examinations is an authentic representation of curriculum aims that include the development of automatic skills and reliable memorisation and recall of facts and procedures. Thus the considered response of the British Secretary of State for Education to the inclusion of an open-ended investigational task as part of national assessment of 14-year-olds in 1991 was to deny its authenticity, describing it as “elaborate nonsense” (Broadfoot & Gipps, 1996).

The reformers’ appeal to authenticity, then, is not an appeal to a newly discovered principle that assessment should match the values of the discipline, curriculum and pedagogy but a move within the ongoing struggle over the values of the curriculum. Within the context of mathematics education, the word has come to be seen as part of a wider ranging discourse of reform (strongly influenced by the United States NCTM). The forms of curriculum, teaching and assessment advocated by this discourse are generally consistent with a ‘competence’ model of pedagogy¹ (Bernstein, 1990) which:

places value on pupils and their internal processes rather than on their finished product; [...] *orientates* teachers to look for evidence of pupils’ progress and development, i.e., what is *present* in the work; and [...] suggests *criteria* of assessment which are *multiple* and *diffuse*. (Morgan, Tsatsaroni, & Lerman, forthcoming)

This is in contrast to forms of curriculum, teaching and assessment within what Bernstein terms a ‘performance’ model of pedagogy. Such forms place value on students’ products (i.e. answers) rather than the processes they may have gone through and orient teachers to look for what is *absent* from the work produced. The criteria of assess-

ment are explicit and can be applied unambiguously. Reform discourses, such as that of the NCTM Standards, generally legitimise themselves by establishing themselves in opposition to a performance model of pedagogy.

The activity of assessment and the associated criteria, whether explicit or implicit are an essential part of any educational practice. Indeed, Bernstein's theory of pedagogic discourse identifies the evaluative rules as constitutive of pedagogic practice — without them there would be no transmission or acquisition but only conversation. The differences in assessment between traditional and reform pedagogies are not only related to the types of tasks used for assessment purposes but also to their orientation and the types of criteria they involve. It is this aspect of authentic assessment practices — the nature and application of their criteria — that I intend to explore in this paper.

Criteria: traditional and reform

In their characterisation of modes of pedagogy found within mathematics education, Lerman & Tsatsaroni (1998) identify the internationally dominant mode as 'traditional'. This is a performance model pedagogic mode, in which the criteria of evaluation are explicit and evaluation processes are orientated towards what is absent from the texts produced by students. This is the mode that is probably most familiar to mathematics teachers around the world as may be seen in the predominance of forms of assessment involving relatively short responses to closed questions given in time-restricted conditions (Clarke, 1996). Assessment tasks are set which are unambiguous in their demands, allowing little doubt about whether a student response is correct. For example, an answer to an examination question will be compared to an 'ideal' answer or marking scheme and, where it differs from the ideal, will be judged to be unsuccessful. Such explicitly stated criteria are, at least in principle, accessible to students as well as assessors; it is clear that all questions have correct answers that are to be achieved by using the correct methods. The correct methods are clearly identified by textbooks, syllabuses, past examination papers, etc. Any ambiguities in the assessment process may, at least in principle, be eliminated by further explicit specification of the criteria.

On the other hand, the competence model pedagogies advocated by many current reforms may be characterised as 'liberal progressive' (Lerman & Tsatsaroni, 1998). They tend to be based theoretically (explicitly or implicitly) on 'child-centred' views of learning. For example, the various versions of constructivism that lie behind many of the assumptions of the *Standards* reform in the USA (National Council of Teachers of Mathematics, 1989, 2000) see learning as a process that depends fundamentally on the individual student's history of knowledge construction. The assessment practices identified as part of this reform pedagogy, either officially sponsored by large scale curriculum developments (e.g. National Council of Teachers of Mathematics, 1995; 2000; Stenmark, 1991) or developed at a more local level by teachers and researchers sympathetic to the reforms, focus on what is present in students and make use of criteria that may be implicit or tacit. The focus on what is present in students allows space for acknowledgement of the historically situated nature of students' construction of knowledge; rather

than measuring all students' performance against a single pre-determined model of correct mathematics, the evaluator must search each student's production individually to identify signs of the mathematical competences they have achieved. The need to consider each student's production on its own merits means that it is much more difficult to specify unambiguous universal criteria. Thus the criteria in such a pedagogic mode are either stated in terms that require expert knowledge to operationalise or are not stated at all and are used only implicitly. In order to explore the implications of such an orientation, in the next section I shall consider the example of 'open-ended' problems (that is, problems that allow a number of different approaches to their successful solution).

Criteria for assessing 'open-ended' problems

Various kinds of open-ended problems have been widely promoted as valuable for use in teaching (see Pehkonen, 1997 for international examples) and have been incorporated into associated reformed assessment regimes, including the GCSE (General Certificate of Education) undertaken at age 16+ by almost all students in England and Wales. Within the GCSE system, introduced in 1988, students' reports of their work on extended open-ended investigative tasks are assessed by their teachers and the results, together with those of more traditional timed examinations, contribute to a final grade that is widely used as a qualification for entry into further education or employment. The use of open-ended problems for assessment purposes is advocated for at least two reasons that arise from a liberal progressive mode of pedagogy. First, it is argued that such problems allow wider access for students to address them with their personal knowledge and experience — demonstrating a child-centred view of knowledge. Second, they enable assessors "to see student thinking rather than test-writer thinking" (Stenmark, 1991, p. 20) — demonstrating an orientation towards what is present in the student rather than on what may be absent from their response.

Let us consider the nature of the criteria used to assess such questions and the extent to which these criteria may be explicit to students. Stenmark suggests that the assessment may be either holistic or analytic. A proposed method for holistic assessment is described as follows:

... first sort the papers into several stacks, representing top-level papers, middle level, and bottom level. If desired, the stacks can be reviewed and re-sorted to form additional categories. (p. 24)

Such a method of assessment, with its obvious lack of explicit criteria, has been criticised as a source of unfairness because of its subjectivity. The teacher or other assessor uses their personal implicit criteria and there is no objective standard against which to measure whether the assessment is 'correct'. Yet in some curriculum areas, at some times, such holistic assessment has been accepted and widely practised. For example, in the 1970s when child-centred pedagogies were influential in the UK and USA holistic assessment of essays produced in language studies was widespread. At this time, Cooper

argued that “holistic evaluation by a human respondent gets us closer to what is essential in such a communication” and reported that rank ordering by experienced assessors could achieve as much as 0.90 reliability (Cooper, 1977). Similarly in mathematics, informal reports of assessment of open-ended tasks suggest that a high degree of agreement among teacher-assessors can be achieved (see, for example, Wiliam, 1994). Nevertheless, the holistic approach, with its acceptance of reliance on tacit criteria, available only to experts, is not in widespread use in mathematics.

The alternative, analytic, assessment involves using a rubric of stated criteria which may be generally applicable or may be designed for a specific task. To consider the nature of the criteria it may be helpful to consider an example of an open-ended problem taken from Stenmark (1991, p. 23):

James knows that half the students from his school are accepted at the public university nearby. Also, half are accepted at the local private college. James thinks that this adds up to 100%, so he will surely be accepted at one or the other institution. Explain why James is wrong. If possible, use a diagram in your explanation.

The rubric for this question was developed by teachers reviewing a set of student responses. Out of the total of 6 possible marks available, a student would be allocated 4 marks if their response “is generally correct, but the explanation lacks clarity”. For 5 marks:

The response is correct and the explanation is clear. It may be expressed in words, with a diagram, or both.

And for the full 6 marks:

The response is exemplary. It goes beyond the criteria for 5 points. For example, the response may include:

- Example(s) and/or counterexample(s)
- Mathematics expressed elegantly
- An explanation that is complete

Correctness is a criterion that is very familiar to mathematics teachers, drawing on the resources of more traditional forms of pedagogy and assessment. It appears to be definable in objective mathematical terms. In this case, however, it is still in need of interpretation. For example, a student response might indicate the construction of an incorrect initial model of the problem but then proceed to use mathematical techniques correctly to produce a solution that is correct in terms of that initial model. Can this be considered worthy of 4 marks — or how many? Even if this ambiguity can be resolved through a more detailed marking scheme (as is often the case for traditional closed questions where, for example, a certain number of marks may be allocated for correct methods even when

these result in wrong answers arising from ‘follow-through’ of errors made early in a solution), other aspects of the rubric, in particular the notions of ‘clarity’ and ‘elegance’ and even ‘completeness’, are, in principle, not amenable to explicit elaboration.

Mathematical elegance, for example, is a concept that only has meaning for those who are ‘insiders’ in particular forms of mathematical discourse. It may be possible to identify some characteristics that are commonly associated with elegant expression, perhaps including brevity, sparing use of symbol manipulation, a high level of generality and generalisability, making use of pre-existing solutions in analogous mathematical domains, but even these characteristics are unlikely to be either necessary or sufficient to distinguish elegance reliably for all mathematicians, mathematics teachers and assessors. Similarly, to say that an explanation has clarity is only meaningful relative to the uses to be made of it in a particular context. For whom is the explanation to be clear — the teacher, the expert mathematician, a fellow student? And for what purpose — to introduce the reader to a new idea, to justify a claim, to display a particular form of understanding, to enable a fellow student to achieve similar results? Each of these audiences and purposes would be likely to demand different amounts of detail about the various aspects of the solution and different forms of language to communicate effectively. (The linguistic characteristics of texts that may be judged to be mathematical, elegant or clear are discussed in more detail in Morgan, 2001).

The experience of holistic assessment of essays mentioned above and the development of assessment of open-ended mathematical investigative work (William, 1994) and of mathematical processes (Roper & MacNamara, 1993) suggests that groups of assessors can achieve a high degree of agreement about the application of such criteria, even though they may not be able to define the characteristics they are looking for. As William states with reference to the development of reliability in assessment of investigative work in the UK, “To put it crudely, it is not necessary for the raters (or anybody else) to know what they are doing, only that they do it right” (p. 60). The rubric Stenmark provides for the question above may in fact be successful in achieving agreement among various mathematics teachers about the marks to be allocated to a given response — particularly where those teachers are drawing on similar backgrounds of experience and training. Mathematics teachers working together with common curricula, text books, professional literature, and common forms of professional development, will share many resources, including curricular values and expectations about student performance, that make it likely that they will respond in broadly similar ways to student texts, even when no criteria are articulated. Nevertheless, the introduction of forms of assessment oriented towards what is present in students has been accompanied by considerable concern for developing reliable scoring methods — drawing on more traditional pedagogic discourse in which the subject matter being assessed is conceived of as identical for all students.

Students and criteria

Much attention has thus been paid to the development of criteria and of teachers' assessment skills (for a fuller review, see Morgan & Watson, 2002). However, we need also to consider the student experience of assessment. In order to be judged to be a successful learner, a student has to know what kinds of performance are recognised as legitimate within the particular school mathematics discourse and are thus likely to be valued by his or her teacher (or other assessor) and must also have the knowledge and skills needed to produce such a performance; in other words, the student must acquire both the recognition rules and the realisation rules of the discourse (Bernstein, 1996). In the case of traditional forms of pedagogy, the explicit nature of assessment criteria makes at least the recognition rules open to inspection by students. Approved answers to exercises are printed at the back of the textbook; teachers model the forms in which they expect students to give their own answers; model answers to past examination questions are provided for students to follow. Of course, this does not necessarily mean that students have access to the realisation rules that would enable them to produce answers that are correct and in the required form. For example, a recent study of English secondary school students' conceptions of mathematical proof found that many students were able to identify the type of argument that their teacher would value most highly, even when they themselves produced arguments in very different forms (Healy & Hoyles, 2000). These students demonstrated awareness of the accepted forms of this aspect of mathematical discourse without necessarily being able to participate in it fully.

In forms of pedagogy in which the criteria are implicit, however, the task of identifying what sorts of performance will be valued is much more difficult. In a study of secondary school students' writing of reports on their investigative work and the ways in which teachers assessed the reports (Morgan, 1998), I found that the type of advice that teachers were offering to their students to help them meet the criteria was not sufficient to help some students to recognise what was required. To illustrate this, I shall consider the case of criteria relating to communication. The criteria provided for the teachers to use were as follows:

- At grade F: Produces some sketches and graphs and, where appropriate, computer output. Able to make limited use of mathematical terms.
- At grade C: Uses an adequate range of mathematical language and symbols, including appropriate visual forms and, where appropriate, computer output. Uses some mathematical words relevant to the task and is generally familiar with the vocabulary of Level 1.
- At grade A: Where appropriate, makes use of symbols when generalising. Selects the most appropriate methods for communicating results. Makes effective use of a range of mathematical language and notation, diagrams, charts and, where appropriate, computer output.

(ULEAC, 1993)

Like the notions of completeness and elegance discussed above, words such as *appropriate*, *effective* and *adequate* used in these criteria may be understood in fairly consistent ways among mathematics teachers who are members of an established community of practice. Making that understanding explicit for students is, however, problematic. A common way in which teachers interpret these official criteria in an attempt to help their students is to advise them to include diagrams in their reports. When students attempt to follow this advice, however, it is not guaranteed that their use of diagrams will be deemed to be 'appropriate'. Indeed, some produce diagrams that actually appear to lower teachers' evaluation of their mathematical performance. One girl, Sandra, for example, working on an investigative problem that involved generating some empirical data in order to observe patterns and form a generalisation, included diagrams in her report that recorded the many empirical examples she had generated (Figure 1). These diagrams were realistic in their representation of the practical equipment she had used (Cuisenaire rods piled on top of one another until the pile toppled over), presenting a three dimensional appearance and even coloured realistically.

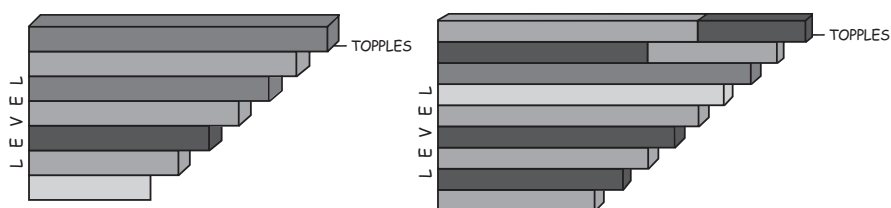


Figure 1. Two of Sandra's many diagrams

The teachers whom I interviewed as they assessed this report were disparaging in their comments about these diagrams, dismissing them as "nice" but "not really necessary", implying that she had wasted time on producing the diagrams that could have been better used on other aspects of the work (although there was no limit to the time students were allowed to work on this task). Moreover, some teachers' subsequent comments on Sandra's work appeared to assume that she was working at a more concrete, practical level than the other students whose work they were assessing, although there was no other evidence of this in her report. It seemed as if the realistic diagrammatic representation of her initial experimentation created such a strong impression for the teachers that they interpreted the rest of her work as being the product of practical, concrete thinking rather than abstract engagement with the task.² In conclusion, Sandra was assessed to be working at a lower level and to have achieved less than other students who included fewer or more abstract, non-naturalistic diagrams such as that shown in Figure 2, although they too must have undertaken practical activity in order to gather their original data.

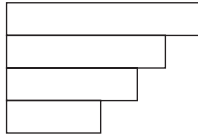


Figure 2: An example of a more abstract diagram

In Sandra's case, the statement of the criteria and her teacher's attempts to interpret the criteria appear to have failed to help her to produce the sort of diagrams that would be valued. Instead, it seems that she has drawn on the resources and criteria of other discourses, with which she was perhaps more familiar, in order to make sense of the advice to draw diagrams. Neatness, accuracy and attractive presentation are highly valued in many school contexts, while displaying work on a large number of similar examples is likely to receive official approval within traditional mathematics pedagogy. Unfortunately for Sandra, these values are not those of the liberal progressive discourse of investigation and open-ended problem solving. The diagrams that she has produced not only fail to be read as evidence of fulfilment of the criteria but are actually read as signs of a low level of mathematical thinking. The implicit nature of the criteria makes it significantly more difficult for students to recognise what is required.

It might be argued that making the criteria more explicit would enable students like Sandra to produce work that would be assessed more positively. Perhaps if she were told: "you must draw a diagram like *this* and include it at *this* point in your report, in order to make *this* point" ...? Such explicitness, however, would entirely change the nature of the pedagogic discourse. The value placed on the student's own experience and activity would be replaced by pre-determined models of 'approved' mathematical texts. The orientation during assessment towards identifying what the student herself has done and is capable of (i.e. what is present) would be replaced by identification of where she has failed to produce a text that matches the expected model answer (i.e. an orientation towards absence). The use of open-ended problems for assessment purposes contains an internal tension between the valuing of students' personal competences and the need to validate these in some way in relation to other students and to some apparently objective notion of mathematical achievement.

Understanding student failure and success

Traditional explanations of students' failure to complete assessment tasks satisfactorily have tended to focus on the characteristics of the individual student. He or she lacks mathematical ability, has not studied hard enough, reacts badly to the pressure of examination, and so on. However, it is possible to adopt an alternative perspective that sees success in school mathematics as participation in a particular form of discourse rather than as acquisition of a body of knowledge. Such a perspective allows us to see fail-

ure as a consequence of students' use of the concepts, values and forms of expression of alternative discourses, different from those of the 'target' discourse. Thus, Sandra's failure to participate successfully in the discourse of mathematical investigation may be interpreted as a consequence of her use of the resources of alternative discourses, perhaps those of a more traditional form of school mathematics, rather than as a lack of mathematical ability, knowledge or understanding. The questions raised from this perspective in relation to assessment thus relate to the nature of participation in discourse rather than to the possibility of determining what individuals do or do not know. Crucially, we need to consider students' access to full participation in the discourses of new forms of pedagogy and assessment.

It has long been a matter of concern that some social groups appear to be disadvantaged by particular forms of assessment; for example, a number of studies have shown girls to do less well than boys when tested by multiple choice rather than open response questions (Gipps & Murphy, 1994). The rhetoric of advocates of authentic assessment claims that equity is one of its central principles (National Council of Teachers of Mathematics, 1995) and the child-centred theories of learning underpinning its pedagogy are in harmony with the claim that each child's competence will be valued in its own right. Nevertheless, there are a number of areas of concern. In the United States, some minority community groups have expressed concern that their children may be disadvantaged by new forms of authentic assessment, pointing out that the emphasis on 'realistic' problems is likely to involve cultural contexts that are not equally familiar to students from different communities (Baker & O'Neil, 1994). In England and Wales, the introduction of assessment by 'coursework' (including independent work carried out over an extended period of time) as a form of authentic assessment across all subject areas has been found to further advantage the same groups of students who were already most successful in traditional forms of assessment (Abrams, 1991; Gillborn & Youdell, 1999; Stobart, Elwood, & Quinlan, 1992). Cooper and Dunne's study of students' responses to different types of items on national Key Stage tests (taken by all students at ages 7, 11 and 14) in the UK demonstrates greater differentiation by social class and by gender on items identified as 'realistic' than on those identified as 'esoteric' (Cooper & Dunne, 2000). The extensive use of 'realistic' problems in these tests may be seen to have arisen from a liberal-progressive discourse of relevance to students' experience. While specific criteria are provided for assessing individual questions, general criteria for questions of this type are more difficult to state explicitly as they involve subtle decisions in each case about how much of the 'real' context needs to be used and how much needs to be 'bracketed out'. Cooper and Dunne argue that, in many cases, children's failure to produce adequate responses to such questions did not arise from a lack of mathematical understanding but either from difficulty in choosing appropriately between using 'everyday' knowledge or mathematical knowledge to solve the problems (lacking the recognition rules) or from lack of awareness of the form of the answers they needed to give (lacking the realisation rules). Working-class students, especially working-class girls, appeared particularly disadvantaged by such items.

The finding that groups that are relatively disadvantaged socially — working class

students, those from some minority ethnic communities — are particularly disadvantaged by these attempts at authentic assessment is consistent with the social theories of both Bernstein (1990) and Bourdieu (Bourdieu, Passeron & de Saint Martin, 1994). Both identify the linguistic competences that students bring with them into the school as critical to their likely ability to participate successfully in the practices of the classroom. The linguistic patterns that middle class students³ experience in their home environments are likely to be closer to the types of interaction that take place in classrooms, thus giving these students more straightforward access to the specialised discourses of the mathematics classroom (Brice-Heath, 1982; Zevenbergen, 1998). When it comes to assessment, the implicit nature of evaluation criteria necessitated by 'reform' pedagogic discourse compounds the social class differences. As Bourdieu and Passeron argue, assessors are likely to infuse their application of implicit criteria with their own unstated values, including those that arise from their own cultural and class background.

Class bias is strongest in those tests which throw the examiner onto the implicit diffuse criteria of the traditional art of grading, such as the dissertation or the oral, an occasion for passing total judgements, armed with the unconscious criteria of social perception on total persons. (Bourdieu & Passeron, 1990, p.162)

It may be argued that assessment practices involving procedures such as moderation could help to safeguard against this sort of bias. Even if this were the case, I would suggest that such practices could not eliminate bias, as it is the discourse of school mathematics itself that disadvantages working class children, not just the prejudices of middle class teachers and assessors.

Implicit criteria act to disadvantage working class students not only because their application is infused with middle class values (indeed, explicit criteria equally reflect the dominant discourse) but also because they are difficult to acquire for those students who do not already share the linguistic code underpinning the discourse. Acquiring the criteria of a traditional form of assessment with explicit criteria entails reproducing the consistent patterns of exemplar answers. It is relatively straightforward to learn that a single letter should be given as response to a multiple choice question, that equations generally have numbers as their solutions, that the sum of two fractions is likely also to be a fraction, and most students successfully provide answers to traditional assessment tasks in a recognisably correct form, even if the correct content escapes them. In contrast, acquiring an understanding of what might be recognised as an *appropriate* form of mathematical communication or an *elegant* solution to an open-ended problem entails sharing a specific set of cultural values that cannot be acquired by such a process of reproduction.

What next?

For mathematics educators who are committed to social justice, the question of what form assessment should take is difficult. It is, of course, possible to opt out of the

question completely, arguing that any form of assessment is a tool used to impose and maintain power and is thus inevitably inequitable; we should struggle to transform education and, indeed, society in order to make assessment irrelevant. Tempting though this argument may be, struggles over the nature of education take place within current dominant educational discourses in which assessment plays a central role. In my opinion it is important to engage with these discourses in ways that raise social justice issues, challenging the assumptions on which both traditional and reform discourses are based. The reformers' advocacy of authentic assessment focuses on the aim of making assessment better — providing more accurate measures of achievement, a more genuine and complete representation of the curriculum and of mathematical values. The arguments I have begun to offer in this paper suggest that, rather than looking at how successful a particular form of assessment may be in the task of assessing, it is important to focus on how processes of assessment work in practice and the social effects of these processes.

In showing how the implicit criteria used in authentic forms of assessment may act to disadvantage some groups of students, I am not proposing that traditional forms of assessment are more equitable. In fact, it is clear that traditional forms of assessment have for a long time been used very effectively as part of the machinery for the maintenance of a stratified society. Whatever the form of assessment may be, the task of the critical mathematics educator must be to investigate how various groups of students are advantaged or disadvantaged by it and to understand relationships between the nature of the assessment instruments, criteria and processes and the success or failure of these groups. Understanding success and failure is a necessary prerequisite not only for the development of strategies to support individuals and groups of students but, more importantly, for posing a challenge to the assumptions and hidden consequences of dominant discourses of mathematics education and assessment.

Notas

¹ Bernstein's use of the terms *competence* and *performance* as descriptors of modes of pedagogy is different from some other uses in assessment contexts. The *performance* model places emphasis on the primacy of the subject discipline and the work that students produce—their performance—is evaluated against a pre-determined model of what is required by the discipline. The *competence* model, in contrast, emphasises the students themselves. The work each student produces is seen as evidence of underlying competence within the student and it is this competence that is evaluated rather than the student's performance on a specific task.

² Kress & van Leeuwen (1996) discuss the characteristics of modern scientific diagrams, making an analogy between simplified forms and lack of setting of such diagrams and the objective, impersonal forms of modern scientific writing (p. 180). The use of perspective and colour is not 'scientific' in this sense and may thus be read as representing activity that is not fully 'scientific' or as the work of an author who is child-like.

³ The argument here is stated in terms of social class differences. Similar arguments may be used to understand the disadvantages experienced in school by some other social groups defined, for example, by cultural or racial origin. In these cases, however, class remains an important factor.

References

- Abrams, F. (1991, 15 February). GCSE fails to eliminate inequality. *Times Educational Supplement*.
- Baker, E. L., & O'Neil, H. F. (1994). Performance assessment and equity: a view from the USA. *Assessment in Education*, 1(1), 11–26.
- Bell, A., Burkhardt, H., & Swan, M. (1992). Balanced assessment of mathematical performance. In R. Lesh & S. J. Lamon (Eds.), *Assessment of Authentic Performance in School Mathematics*. Washington D.C.: American Association for the Advancement of Science.
- Bernstein, B. (1990). *Class, Codes and Control, Vol.IV: The Structuring of Pedagogic Discourse*. London: Routledge.
- Bernstein, B. (1996). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. London: Taylor & Francis.
- Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in Education, Society and Culture* (2nd ed.). London: Sage.
- Bourdieu, P., Passeron, J.-C., & de Saint Martin, M. (1994). *Academic Discourse: Linguistic Misunderstanding and Professorial Power* (R. Teese, Trans.). Cambridge: Polity Press.
- Brice-Heath, S. (1982). Questioning at home and school: A comparative study. In G. D. Spindler (Ed.), *Doing the Ethnography of Schooling*. New York: Holt, Rinehart & Winston.
- Broadfoot, P., & Gipps, C. (1996). Assessment developments in England and Wales: The triumph of tradition. In A. Little & A. Wolf (Eds.), *Assessment in Transition: Learning, Monitoring and Selection in International Perspective* (pp. 134–153). Oxford: Pergamon.
- Clarke, D. (1996). Assessment. In Alan J. Bishop, Ken Clements, Christine Keitel, Jeremy Kilpatrick, and Colette Laborde (Eds.), *International Handbook of Mathematics Education* (Vol. 1, pp. 327–370). Dordrecht: Kluwer.
- Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper & L. Odell (Eds.), *Evaluating Writing: Describing, measuring, judging* (pp. 3–31). New York: National Council of Teachers of English.
- Cooper, B., & Dunne, M. (2000). *Assessing Children's Mathematical Knowledge: Social Class, Sex and Problem-Solving*. Buckingham: Open University Press.
- Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education*, 6(2), 177–194.
- Gillborn, D., & Youdell, D. (1999, 26 November). Weakest not at the table. *Times Educational Supplement*, 13.
- Gipps, C., & Murphy, P. (1994). *A Fair Test? Assessment, Achievement and Equity*. Buckingham: Open University Press.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31(4), 396–428.
- Kress, G., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Lerman, S., & Tsatsaroni, A. (1998). Why children fail and what the field of mathematics education can do about it: The role of sociology. In P. Gates (Ed.), *Proceedings of the First International Mathematics Education and Society Conference* (pp. 26–33). Nottingham: Centre for the Study of Mathematics Education, Nottingham University.
- Morgan, C. (1998). *Writing Mathematically: The Discourse of Investigation*. London: Falmer.
- Morgan, C. (2001). Mathematics and human activity: representation in mathematical writing. In C. Morgan & K. Jones (Eds.), *Research in Mathematics Education Volume 3: Papers of the British Society for Research into Learning Mathematics* (pp. 169–182). London: British Society for Research into Learning Mathematics.
- Morgan, C., Tsatsaroni, A., & Lerman, S. (forthcoming). Mathematics teachers' positions and practices in discourses of assessment. *British Journal of Sociology of Education*.
- Morgan, C., & Watson, A. (2002). The interpretative nature of teachers' assessment of pupils' mathematics: issues for equity. *Journal for Research in Mathematics Education*, 33(2), 78–110.

- National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (1995). *Assessment Standards for School Mathematics*. Reston VA: National Council of Teachers of Mathematics.
- National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM.
- Niss, M. (Ed.). (1993a). *Cases of Assessment in Mathematics Education: An ICMI Study*. Dordrecht: Kluwer Academic Publishers.
- Niss, M. (Ed.). (1993b). *Investigations into Assessment in Mathematics Education: An ICMI Study*. Dordrecht: Kluwer Academic Publishers.
- Pehkonen, E. (Ed.). (1997). *Using Open-Ended Problems in Mathematics*. Helsinki: Helsinki University, Department of Teacher Education.
- Romberg, T. A. (Ed.). (1995). *Reform in School Mathematics and Authentic Assessment*. New York: SUNY Press.
- Roper, T., & MacNamara, A. (1993). Teacher assessment of mathematics Attainment Target 1 (MA1). *British Journal of Curriculum and Assessment*, 4(1), 16–19.
- Stenmark, J. K. (Ed.). (1991). *Mathematics Assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: National Council of Teachers of Mathematics.
- Stobart, G., Elwood, J., & Quinlan, M. (1992). Gender bias in examinations: How equal are the opportunities? *British Educational Research Journal*, 18(3), 261–176.
- Torrance, H. (Ed.). (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.
- ULEAC. (1993). *Coursework Assessment 1993: Mathematics A*. London: University of London Examinations and Assessment Council.
- William, D. (1994). Assessing authentic tasks: Alternatives to mark-schemes. *Nordic Studies in Mathematics Education*, 2(1), 48–68.
- Zevenbergen, R. (1998). Classroom interactions and linguistic capital: A Bourdeuian analysis of the construction of difference in mathematics education. In P. Gates (Ed.), *Mathematics Education and Society: Proceedings of the First International Mathematics Education and Society Conference* (pp. 360–366). Nottingham: Centre for the Study of Mathematics Education, University of Nottingham.

Resumo. Novas orientações curriculares para a Matemática fazem frequentemente apelo ao recurso a formas de avaliação mais autênticas. Este apelo é geralmente associado à intenção de aumentar a acessibilidade da educação matemática e das competências matemáticas a todos os alunos. Este artigo apresenta uma perspectiva crítica sobre esta noção. Tomando por base a teoria de Bernstein sobre o discurso pedagógico, argumenta-se que formas de avaliação defendidas por pedagogias “inovadoras” envolvem o recurso a critérios que ou são implícitos ou requerem um conhecimento de especialista para os operacionalizar. A dependência nestes tipos de critérios pode ser desvantajoso para alunos de certos grupos sociais.

Palavras-chave: Critérios de avaliação; avaliação autêntica; Bernstein's model of pedagogy; equidade; tarefas abertas; discurso pedagógico; justiça social.

Abstract. Advocates of reform of mathematics curricula frequently call for the introduction of more authentic forms of assessment. This is often associated with an intention to increase the accessibility of mathematics education and qualifications for all students. This article develops a critical perspective on this notion. Drawing on Bernstein's theory of pedagogic discourse, it is

argued that forms of assessment advocated by 'reform' pedagogies involve the use of criteria that are either implicit or require expert knowledge to operationalise. Dependence on such criteria is likely to disadvantage students from certain social groups.

Keywords: Assessment criteria; authentic assessment; Bernstein's model of pedagogy; equity; open-ended tasks; pedagogic discourses; social justice.

■ ■ ■

CANDIA MORGAN
Institute of Education, University of London
UK
c_morgan@ioe.ac.uk