‘BETTER’ ASSESSMENT IN MATHEMATICS EDUCATION: A SOCIAL PERSPECTIVE

Candia Morgan
Institute of Education, University of London

Introduction

Assessment of school mathematics has an important impact both on individual students and teachers and on groups within society. Its summative forms have traditionally been used in many countries as a primary means of discriminating between individuals in order to allocate scarce resources (often in the form of further educational and employment opportunities). It may be used to compare different groups within a population and thus has potential power as a tool of systematic discrimination, for example, using the results of tests to make and justify policy decisions about the forms of education that should be made available to various groups within a society. At various times and places, the results achieved by their students have also influenced teachers' pay and possibilities for advancement. Even at the classroom level, the day-to-day judgements of teachers about individual pupils inevitably affect future interactions, judgements and hence opportunities. The act of assessment takes place in interaction between individual students, teachers and social contexts and its results have far reaching social consequences, yet research within mathematics education that acknowledges the importance of a social perspective has overwhelmingly focused on curriculum and classrooms rather than on assessment.

The vast majority of existing research related to assessment within mathematics education has been done within two main traditions that I shall characterise as 'psychological' and 'curriculum reform'. Although there is a body of research that addresses assessment from a sociological perspective, this has not yet substantially informed research within the mathematics education community. In this chapter, I intend to critique this 'mainstream' mathematics assessment research and to discuss what might be meant by a social perspective on assessment and research into assessment in mathematics education. I shall also discuss some of the issues that arise, referring to examples of research that have taken or begun to take such a perspective. In conclusion, I shall suggest a number of questions that may be used to interrogate existing research on assessment in mathematics education and to guide future efforts.

Mainstream Traditions of Mathematics Assessment Research.

There are three basic assumptions underpinning mainstream thinking on assessment. First, it is assumed that individuals possess attributes (such as knowledge, understanding, skill, ability, etc.) that are discoverable and measurable. Second, the primary purpose of assessment is seen to be to discover and measure these attributes. Finally, the assessment process and its outcomes are assumed to be fundamentally benign or even beneficial (although

unfortunate side-effects may be recognised and attempts made to ameliorate them). The benefits that 'good' assessment is claimed to bring include: increased understanding of students' mathematical thinking, both collectively and individually; possibilities for improved teaching, with the design of schemes of work that match what is known about students; and opportunities to influence teachers' practice, bringing it in line with the ideals of those designing the assessment.

It is clear that these assumptions are rooted in a strongly positivist tradition. That is, they are predicated on the belief that there is an underlying truth to be assessed and that it is theoretically possible to get as close as one might wish to this underlying truth. This positivist tradition is perhaps even stronger in mathematics than in other subject areas: as many believe that there are only right or wrong answers; people either know the right answer or they do not. Uncertainty and non-excluded middles in mathematical contexts are deeply uncomfortable for many people, even for those who might find them less surprising in other disciplines. Situated theories of learning challenge both the idea that there is some absolute 'truth' about students' understanding of mathematics and the idea that any instrument could observe and measure such a state. Yet such epistemological concerns have had little impact on thinking about assessment (Galbraith, 1993).

The psychological tradition

Research related to assessment that is reported in the Proceedings of PME[1] (the International Group for the Psychology of Mathematics Education) has two main concerns. One is the development and use of assessment instruments to form improved (more valid) characterisations of the attributes of individual students. (See, for example, Hunter & Monaghan, 1996; Jaime & Gutiérrez, 1994.; Karsenty & Vinner, 1996; Lawrie, 1998; Leung, 1994; Rickards & Fisher, 1997) The second is the construction of models of the general characteristics of knowledge and understanding in a given area of mathematical activity (e.g., Garcia-Cruz & Martinón, 1997; Garcia-Cruz & Martinón, 1998; Orton & Orton, 1996). The aspects of mathematics involved include both 'traditional' areas of study, such as 'geometry', and areas associated with current curriculum reform movements, such as 'problem solving'. Some of the studies reported appear to be 'pure' research while others aim to provide tools for teachers to use or to influence teachers' practice, often linked to the 'reform' tradition (see below). While there are substantial differences in the aims, content and theoretical framing of these studies, they all share the three assumptions outlined above. In particular, they strive towards validity in order to discover the truth about students' knowledge - a completely appropriate aim for research in a cognitive psychological paradigm.

---

[1] I believe I am justified in taking the research reported in PME to be representative of the psychological tradition in mathematics education research. There are, of course, other places where work in this tradition is published. My purpose here is not, however, to review the field but to characterise it.

In recent years, learning has increasingly come to be seen as a social rather than solely individual process. Björkqvist (1997) has argued that this means mathematics education assessment research needs to take more account of social psychological frameworks. However, this does not yet seem to have had a significant influence. Moreover, Björkvist's primary concern, like that of others within the psychological tradition, is still with the search for validity. There is little room for considering either the social effects of the assessment process or the possibility that validity might be an illusory goal.

The curriculum reform tradition

The field of mathematics education has witnessed many arguments about both the nature of the curriculum and the nature of its assessment. The link between curriculum and assessment is now explicitly seen as a two-way relationship and there has been increasing recognition of the role that assessment structures play in influencing what happens in classrooms, either hampering or 'leading' efforts at curriculum reform (Burkhardt, 1988; Ridgway & Schoenfeld, 1994). The function of assessment that is of primary importance for curriculum reformers is the power to influence the curriculum, with accompanying concern that this influence should 'match' the intentions of the reform. The focus of the curriculum reform tradition in relation to assessment has thus been two-fold. First, researchers in this tradition have critiqued existing forms of assessment, considering both their effects on the curriculum and their effectiveness in producing valid measures of those aspects of students' mathematical thinking that are valued by the proposed curriculum. Second, they have proposed, developed and trialled alternative forms of assessment intended to match more closely the aims of the reform curriculum, both in producing the desired effects in schools and classrooms, and in measuring that which the reform values.

In particular, there has been increasing interest in the idea of 'authentic' assessment, in mathematics as in other subject areas (Torrance, 1995). The term 'authentic' has a number of connotations, not all of which are made explicit in debates about assessment reform. Authentic tasks may have some 'realistic' aspect; their assessment procedures may match more closely the aims, content and breadth of the curriculum; and the results of the assessment may be a more genuine measure of students' achievement. There has been considerable international interest in discussing the characteristics of more 'authentic' means of assessment as well as attempts to design and implement new forms of assessment in the context of curriculum reform (Leder, 1992; Lesh & Lamon, 1992; Niss, 1993a; Niss, 1993b; Romberg, 1995).

Concern for social issues, in particular for 'equity' in relation to gender and minority ethnic groups, has been one of the motivations of recent curriculum reform movements and this has been reflected in some of the principles underpinning parallel assessment reforms. In the United States, for example, the 'Assessment Standards' produced by the National Council of Teachers of Mathematics include the principle that assessment should 'promote equity' (NCTM, 1995, p.15). The dissatisfaction with traditional forms of assessment, in particular with timed tests demanding pre-determined short answers to closed questions has arisen in part from a growing awareness of cultural bias in even (or perhaps especially) the most 'objective' of tests.

The power of assessment to influence the curriculum is, however, a double-edged sword. It is necessary to ask who is controlling the reform and in whose interests they act. In recent years in the United Kingdom, we have seen a change in the relationships between teachers, curriculum reforms and assessment practices. In the 1970s and early 1980s, the official examination system allowed some possibilities for teachers to develop alternative practices. Although regional examination boards controlled the setting and marking of high-stakes examinations for students aged 16+, schools and groups of schools were able to apply for validation of their own syllabuses and examinations, while independent curriculum developers such as SMP, SMILE[2] and the Association of Teachers of Mathematics were successful in developing alternative syllabuses and assessment regimes that could be adopted by schools. During this period, reformers who wished to see greater diversity in the curriculum and opportunities for wider groups of students to participate in mathematics made use of innovative assessment methods to encourage the teaching of problem solving and the use of mathematical investigation in the classroom (see, for example, Love, 1981). In 1988, with the introduction of a new national system of examination for England and Wales, some of these practices were officially endorsed and, eventually, made compulsory. As I have argued elsewhere (Morgan, 1998b), this use of assessment to instigate universal reform actually acted in some cases to distort and impoverish the types of rich mathematical activity it was intended to encourage. Since the late 1980s, assessment has increasingly been used as a tool in the move towards centralised control of the curriculum. Teachers have lost most of their opportunities to innovate and to have their innovations validated through the official assessment system. Both the content and the method of teaching have been deliberately engineered through the introduction and shaping of national tests for political as much as educational purposes. As Galbraith argues, the generally accepted idea that external assessment requirements should be used to influence the curriculum is "ultimately disempowering to teachers in impeding the growth of full professional responsibility, and to students in making their choices and interests irrelevant." (Galbraith, 1993, p.82).

Critique of the mainstream traditions.

Both psychological and curriculum reform traditions of assessment research have as their principle aim the search for validity, in the narrow sense of accuracy of measurement (of whatever is being measured) and in the broader senses of curricular and content validity. Their efforts are thus directed to the development of new and better assessment instruments. Critiques of traditional methods of assessment still tend to lie within the same positivist paradigm. For example, in the recent ICMI study on "Mathematics Education as a Research Domain" Ellerton and Clements (1998)

---

[2] SMP (School Mathematics Project) involved academics and teachers (initially largely from independent schools) in developing text books. The materials are very widely used in schools throughout England. SMILE (Secondary Mathematics Individualised Learning Experiment – later changed to 'Experience') was developed by groups of teachers, mainly within London, who produced materials that allowed students to work on individually designed programmes. Teacher involvement in developing materials continues, though recent years have seen a decline in the number of schools adopting the individualised way of working.

identify "Questioning the Basis for Assessing Achievement in Mathematics" (p. 168) as one of ten major concerns for mathematics education research. Yet their 'questioning' leads primarily to the claim that "Closer research scrutiny needs to be given to the issue of how achievement is best measured in mathematics" (p. 169). Their point is illustrated by reference to research by Thongtawat which "found that students who scored poorly on a [multiple choice] test could sometimes have a good conceptual grasp of the material which the items covered" (p. 169). This critique assumes that some (unspecified) alternative method of assessment can provide a more valid picture of students' 'conceptual grasp'. Even Galbraith's (1993) critique of the inconsistency of those who simultaneously espouse constructivist epistemologies and cling to positivist conceptions of assessment is not consistent. He too appears to argue that at least 'local' forms of assessment that take into account cultural and contextual features could provide valid knowledge of students' understanding. He rejects the positivist principles of reliability and objectivity but still relies on the idea that it is possible to seek validity.

Although it is generally recognised that students interpret what teachers say in multiple ways, this insight into the contingent nature of meaning making is rarely extended to how teachers interpret what students say or write. Mainstream thinking about assessment is still based on a 'common sense' or naive transmission view of the nature of communication in which meaning resides within the text, independent of the reader, carrying the author's intentions exactly. The assessor's role is thus to 'extract the meaning' from the text produced by the student. Obvious failures to communicate - where different modes of communication (for example, a written test and a teacher observation of a child working) provide different messages about the 'same' student competence or where the teacher/assessor is unable to make sense of a written or spoken text produced by a student - are often seen to be a 'language problem' for the student. But on what basis do we assume that, when teachers and other assessors do succeed in making sense of a student's text, they then know what the student intended to communicate? A more consistent epistemology would suggest that there is no necessary simple correspondence between a piece of text and the meanings its various readers construct. Rather, the meanings constructed will depend on the resources brought to bear on the text by individual readers. These resources will vary according to the discourse within which the text is read and the positions adopted by a particular reader within that discourse as well as the reader's previous experience (Kress, 1989). There can never be a guarantee that the interpretations made by the assessor are exactly those intended by the student. Indeed, studies of teacher/assessors demonstrate how different assessors can construct entirely different interpretations from the same text (Morgan, 1996; Morgan & Watson, forthcoming).

Towards a Social Perspective on Assessment in Mathematics Education

I have argued above that the aim of mainstream research related to assessment in mathematics education has generally been to develop and validate more 'accurate' and 'authentic' means of assessing student learning. I have also argued that this aim is an illusory one - that the first two assumptions on which both psychological and curriculum reform traditions are based are unfounded and inconsistent with epistemological stances that take into account the socially situated and contingent nature of knowledge and of communication. I now turn to the third assumption of the

mainstream traditions - the assumption that assessment is fundamentally benign. In making this assumption, it is clear that mainstream research neglects the crucial functions that assessment performs within society.

What is a 'social perspective'?

What might constitute research with a 'social perspective' on assessment in mathematics education? The term 'social' can have reference at a number of different levels from interactions between two individuals to national and international structures and systems. But whatever aspect of the 'social' is the object of study, I would argue that such a perspective must involve some consideration of values and consequences. Thus it must involve asking the questions:

- who benefits and who is disadvantaged?
- how do assessment processes and systems act to benefit or disadvantage individuals and groups?

In each case, the scope of the questions may vary from the level of an individual classroom to large scale assessment systems, both national and international. Examples of questions at each of those levels could be:

- Who within this classroom is disadvantaged by the assessment that takes place?
- How does the form of these national examinations serve to disadvantage working class children?
- How are the opportunities for learning in this country influenced by their government's desire to demonstrate international comparability of qualifications or to raise the country's standing in international comparisons such as TIMSS?

Unlike the mainstream tradition of assessment research discussed above, the object of research on assessment taking a social perspective cannot be the development of 'more valid' assessment instruments, though lack of validity may be one aspect of the critique of existing instruments.

Assessment as discrimination

Although I have argued that the mainstream tradition assumes that the primary purpose of assessment in mathematics is to discover and measure the mathematical attributes of individuals, it is, of course, generally acknowledged that the results may be used to serve a wide variety of functions, ranging from the design of individual learning programmes, to determining students' future educational or occupational opportunities, to evaluating teachers. All these functions, however, are predicated on the belief that knowing the 'truth' about an individual's state of knowledge or understanding is a proper basis for making such decisions. As Broadfoot (1996) points out, this belief is a relatively modern development, reflecting the dominance in modern society of individualism and of rational (rather than coercive) authority as the basis for hierarchical control. It is no longer acceptable, as it once was, to say that a certain type of

education is not suitable for a particular group of children because their parents are peasants or factory workers. It is similarly unacceptable to say that this person is the right person to become a doctor or a lawyer or a priest because he comes from the appropriate social group. However, the dominant modern ideology allows arguments of the form: this child should be taught in this way because of her own talents or difficulties; this person should be trained to be a lawyer because she has demonstrated that she has the necessary intellectual and personal attributes (including the desire to become a lawyer). Whether rational authority and the resulting 'meritocracy' is superior (ethically or practically) to other bases of discriminating between people in order to organise the division of labour and distribution of resources within society is not an issue I wish to address here. The point is that this ideology is a means of legitimating divisions and control within society.

Rather than seeing assessment as the attempt to discover the truth of a student's state of understanding, it is thus useful to look at it as the process by which a student may gain or be denied access to particular forms of privilege or power. We know that mathematics qualifications serve in many societies around the world as a means of discriminating between individuals when allocating educational and occupational opportunities, even where knowledge of mathematics itself may be irrelevant to the future performance of the individual. As Noss claimed in his critique of the UK National Curriculum, the purpose of assessing ability to perform long division is to "divide and rule" (Noss, 1990).

There is a tension for mathematics educators in taking such a perspective on their activity as most have entered the field with a major interest in mathematics and in students' learning. If schooling and assessment are regarded as a means of discriminating between individuals then the mathematics and the learning become unimportant by-product. Vinner (1997) takes a moral stance against the pursuit (by teachers and students) of 'pseudo-knowledge' (that is, knowledge that enables acceptable answers to be given to school mathematics questions but does not represent ‘true’ mathematical understanding) as a means of gaining 'credit' within the system. He claims that the 'intention' of the educational system is 'to give credit for true knowledge and not for pseudo-knowledge' (p.68) and that to gain credit for pseudo-knowledge is a form of 'cheating'. It is not clear to me how a system can have 'intentions', though individuals and groups within it can. What is clear is that the system functions to discriminate and select on the basis of the credits gained rather than the quality of the knowledge acquired.[3]

---

[3] Baldino (1998) proposes a solution to the mismatch in the mathematics classroom between discourse (which claims to value learning) and action (directed towards gaining credit) that leads to what he terms 'cynical consciousness'. He suggests that hard work (always a hidden criterion for success) should be made an explicit criterion. Indeed, he has gone so far as to suggest that it should be the only criterion for awarding credit (personal communication). While this might be a possible approach for individual teachers in education systems which allow them substantial autonomy in deciding how to assess their students, it is not a practical proposition for those teachers who are themselves subject to surveillance and evaluation.

What is it that assessment gives credit for? I would agree with Vinner that it is not 'knowledge', because this is not directly observable, but the extent to which the student is able to participate in the form of discourse that is valued within the particular classroom or assessment regime. 'Knowledge' itself will not be given credit unless it somehow manifests itself in ways that are recognised within that discourse. A teacher makes judgements about a student's mathematical knowledge by interpreting oral and written texts and other behaviour produced by the student. The teacher's interpretation of these texts as evidence of knowledge or lack of knowledge depends on the extent to which their form matches the teacher's expectations of mathematical texts – expectations shaped by his or her own knowledge and experience of discourses of mathematics and mathematics education. To be successful in gaining credit, therefore, the student must learn to produce texts that will be judged to be legitimate texts (Bernstein, 1996) within the practices of the mathematics classroom.

It can be argued that learning to participate in mathematical discourse is learning mathematics. I would also argue that the discourse of mathematics classrooms is mathematical - not because it necessarily matches up to some ideal of 'real' mathematical thinking (as defined by mathematicians or mathematics educators) but because it happens in mathematics classrooms (see Morgan, 1998a for a fuller discussion of what may be labelled 'mathematical'). Thus all those who participate to any extent are learning mathematics, even if much of it might be classified by Vinner as pseudo-mathematics. The problem is that much discourse in mathematics classrooms consists of forms of mathematics that are of low status and lack power to enable students either to gain valued credit within the existing society or to take action to transform that society. Dowling's (1998) analysis of the discourse of a scheme of mathematics text books demonstrates how these act to construct and confirm students in their class positions, allowing the 'high attainers' access to high status esoteric domain mathematics that could serve to apprentice them as mathematicians as well as leading to valued qualifications, while the texts for 'low attainers' position them as manual workers and induct them into a narrow and low status form of 'real world' mathematical activity. Dowling does not address assessment directly but his work provides some insight into how division and disadvantage may be constructed in classrooms. The form of the public examination system in the United Kingdom legitimates and necessitates this division. Different groups of students are only allowed to enter one of several 'tiers' of examination, usually determined by their teachers. Only those entered for the highest tier have access to the highest grades while those entered for the lowest tier may achieve a maximum grade of 'D' - largely considered to be a failing grade and carrying little value as a qualification for employment or further education opportunities (see Boaler, 1997, for students' responses to this). (~~?~~Ok – please insert the reference)

Existing Research with a Social Perspective

I turn now to consider the research into assessment of mathematics that has addressed the two questions (who benefits/is disadvantaged and how) suggested above as central to a social perspective. There are three main types of issues that have been addressed:

- differential access to 'high-stakes' assessment and qualifications;
- differential performance on various types of assessment tasks;
- potential inequity in judgements formed by assessors.

I do not aim to present a full review in this chapter but to illustrate some of the ways in which these issues have been addressed and to discuss some of the questions and methodological issues that remain. Most of the research I shall discuss is based within the education system of England and Wales. While I would suggest that the three issues identified above are also likely to be of importance in other education systems, there may well be other issues that arise elsewhere, particularly in 'developing' nations.

Differential access to high-stakes assessment

One obvious way in which individuals and groups can be disadvantaged by an assessment system is through being excluded from access to those forms of assessment that lead to benefits (in the form of high status qualifications or opportunities to progress within the school). As I have described above, the mathematics examination system in England and Wales at age 16+ is structured so that only those considered by their teachers to be 'suitable' may be entered for the tier of examination that allows access to the highest grades - others must make do with second or third rate qualifications. (This is not the case in all subject areas; in English, for example, all students are entered for the same papers.) Stobart et al.'s (1992) study showed that girls are entered in disproportionate numbers for the 'intermediate' tier of examination. Although success in this examination allows students to achieve the 'C' grade that is a requirement for many further courses and employment (including nursing and teaching - traditional women's work) it is not usually accepted as an adequate basis for advanced study in mathematics. The researchers suggest that the decision to enter some girls for the intermediate rather than the higher examination may be due to a perception shared by both teachers and pupils that girls lack confidence in mathematics. (See also discussion of similar decisions about entry to high and low status examination by Walden & Walkerdine, 1985.) Dunne's (1998) study of entries for National Curriculum tests at Key Stage 3 (age 14) also shows that different schools use different bases for making decisions about the level test for which pupils are to be entered. Thus a pupil might be entered for the highest tier in one school, but the middle or lowest tier in another.

The use of results of examinations to evaluate and rank schools encourages schools, departments and teachers to 'play safe' in the decisions they make about entries in order to maximise the appearance of the results of the school as a whole. In the UK, the principle measure used to compare schools is the number of A to C grades achieved in the GCSE examination. As a grade C can be achieved in the intermediate level examination, there is substantial anecdotal evidence to suggest that schools are making more entries at intermediate level because it is perceived that it is easier to gain a C by that route. The interests of the school but not necessarily the interests of the individual student may be served by such a decision. Moreover, it seems that some schools target scarce resources upon students who are borderline between C and D grades, providing extra lessons and support for these students - the motivation being to maximise the school's achievement rather than that of the students (Gillborn & Youdell, 1999).

The provision of differentiated examinations is often justified by suggesting that it allows all candidates to attempt tasks at a level with which they can cope and to achieve some measure of success (Cockcroft, 1982). As soon as decisions must be made about entry levels, however, it becomes clear that some students or groups of students may be disadvantaged by variation in the way the decisions are made and applied to them. The way in which teachers arrive at the assessments on which such decisions are based forms another area of research and is discussed below.

Differential performance on various types of assessment tasks

Because of the prominent place of mathematics in the curriculum and the high status attached to qualifications in mathematics, the subject has often been included as an example within more general studies of bias in testing and examinations. Studies of responses to different forms of assessment tasks have shown that differences between the performance of girls and boys vary between multiple choice tests, 'free response' questions and extended 'coursework' tasks (Anderson, 1989 ; Stobart et al., 1992). These studies and others addressing the idea of 'bias', (see Gipps & Murphy, 1994 for a much fuller discussion) tend to rest on an assumption that it is theoretically possible to make use of the knowledge about differences that they have identified in order to construct assessment instruments that are not biased - thus working with the mainstream assumptions I have critiqued above.

Within mathematics education, the 'context' within which assessment tasks are presented has been seen to be an important variable in making the tasks easier or more difficult for various groups of students. In some cases, this has been based on rather simplistic models of the way in which such bias might operate. For example, a sub-committee of the Mathematical Association (MA) entitled 'Monitoring GCSE[4] Papers for Bias' defined its task to be to examine the contexts of examination questions for their 'cultural diversity', reporting in 1990 that:

---

[4] General Certificate of Secondary Education is an examination taken by nearly all students in the United Kingdom at age 16+.

> ... some papers contain an imbalance in mentions of male/female, and traditional stereotypical roles for girls and women ... There is a little use of 'Asian' names, and a very inappropriate drawing of black children which offends, but no signs at all that any paper acknowledges, yet alone promotes, multiculturalism. (Mathematical Association, 1990)

While the MA report bases its investigation on general principles about a 'multicultural' approach to mathematics education rather than specific claims about the effects of such 'bias' on students' performance in the examinations, it also cites extracts from the policy statements of examination boards which suggest that their concern is underpinned by the assumption that the cultural context does have such effects:

> The Board will be alert to the need to be fair to its candidates by developing schemes of examination whose components do not bias against particular ethnic groups. The culture free examination is an unrealistic aim and would, in any case, lead to dull examinations.
> (London Examinations and Assessment Group, cited in Mathematical Association, 1990)

This is a rather naïve approach to issues of cultural disadvantage. Where the effects of context on students' mathematical problem solving behaviour and success have been studied, it is apparent that the simplistic equation 'culturally familiar/sympathetic context = easier' does not work. Indeed there are indications that familiar contexts can make problems more difficult for some groups of students (Boaler, 1993; 1994; Cooper, 1998; Cooper & Dunne, 1998).

A more theoretically informed approach is taken by Cooper and Dunne in their studies of students' approaches to and performance on assessment tasks (taken from national tests for 11- and 14-year-olds) set in 'everyday' contexts. In comparing the performance of 11-year-old students from working class and professional families, they show that working class children, already achieving at a lower level overall, were even less successful on 'realistic' questions (Cooper & Dunne, 1998). Whereas the rules for answering traditional 'esoteric' mathematics questions are clear-cut, in order to answer such contextualised questions successfully, students have to judge very finely exactly how much everyday 'realistic' knowledge to use. The relatively poor performance of working class children on such contextualised tasks appears to be related to their use of inappropriate 'everyday' modes of response when they would need to draw on more formal mathematical methods in order to achieve the answers expected by the test setters (Cooper, 1998). The researchers draw on the work of Bourdieu (1990) and Bernstein (1996) to explain differences in competence on these tasks between students from different social groups.

Potential inequity in judgements formed by assessors

In rejecting so-called 'objective' multiple choice tests as invalid in relation to the curriculum and inequitable in the opportunities they provide for different groups to display competence, the question of how assessors interpret and form judgements about student responses to assessment tasks becomes crucial. Moreover, teachers' judgements about their students not only may contribute to summative assessments (including the decisions about access to qualifications

discussed above) but also influence their day-to-day interactions and the opportunities they provide their students for learning. (A more detailed discussion of the sources of inequity in teacher assessment may be found in Morgan & Watson, (forthcoming).) It has long been established that teacher expectations and teacher stereotyping of student characteristics can lead to differential treatment of students, to differences in student performance and to differences in the ways in which student behaviour is interpreted and evaluated. At the same time, however, it must be acknowledged that, in most cases, such inequity does not arise from any deliberate discrimination on the part of the teacher/assessor. It is, therefore, important to ask how it does arise and whether, through understanding its nature and sources, it can be challenged.

Studies of teacher assessment in mathematics have suggested that much assessment of students makes use of general constructs of 'ability' or 'level' (see, for example, Gill, 1993; Ruthven, 1987) and that teachers have confidence in the "implicit and largely unarticulated process" (Dunne, 1998, p.153) by which they reach their judgements about student 'ability'. Dunne (1998) points out that these judgements are likely to be influenced by the presence or absence of 'cultural affinity' between teacher and students. Studies of the detail of teacher/assessor interpretation of texts produced by students (in the form of observable behaviour, spoken or written production) (Morgan, 1996; 1998b; Watson, 1997; Watson, 1998) reveal the extent to which different teachers may attend to different aspects and construct very different meanings from the same student text.

The implicit nature of the assessment process identified by Dunne in relation to general judgements of 'ability' also applies to what might appear to be much more concrete cases of assessment of particular incidents or examples of student texts. When studying teachers evaluating student reports of mathematical investigations, I found that, while the teachers were able to rank student texts without difficulty and were able to point to some of those aspects of the texts that they approved or disapproved, they were unable to articulate the evaluation criteria they were using in a way specific enough to help students to produce acceptable texts (Morgan, 1998b). For example, while a teacher might provide general advice to his or her students to make use of diagrams in their reports, when the same teacher was actually reading a text with diagrams in it, only some kinds of diagrams would be valued while others were condemned as 'a waste of time' or apparently taken as a sign of a limited level of mathematical thinking.

The challenge for the student, then, is not to acquire knowledge and understanding of mathematics but to acquire knowledge of the characteristics of the forms of behaviour that will allow her to be seen to know and understand, together with the skills necessary to display the appropriate behaviour. In Bernstein's terms, she needs to acquire the recognition rules that "regulate what meanings are relevant" and the realization rules that "regulate how the meanings are to be put together to create the legitimate text" (Bernstein, 1996, p. 32). The ideals of 'reform' mathematics curricula, unfortunately, increase this challenge for the student. By weakening the framing of the pedagogic discourse - valuing creativity rather than industry, student empowerment rather than rule following - the criteria by which students are to be evaluated become increasingly implicit and invisible. This does not mean that assessment criteria are any less determinate, merely that it more difficult to determine what they are. Lerman and Tsatsaroni (Lerman & Tsatsaroni,

1998) have argued that, just as traditional (strongly framed) forms of pedagogic discourse are inaccessible to working class students, these same students may be further disadvantaged by the discourse of 'reform' curricula and evaluation practices. The implicit evaluation 'rules' are likely to be most accessible to those groups of students whose cultural and linguistic background is closest to that of the school.

Conclusion

In this chapter, I have called into question the assumptions underlying mainstream research into assessment in mathematics education, in particular the search for 'validity', recognising that assessment is an interpersonal interpretative practice, not a scientific measurement, and challenging the assumption that assessment is fundamentally benign. Assessment takes place and acts at a number of different levels: in everyday interactions between individual teachers and students; within classes, departments and schools; in national and international systems. In particular, it functions to discriminate between individuals and groups at all these levels. This discriminatory function needs to be recognised and examined.

Mainstream mathematics education research has taken as its major aim the search for 'better' assessment instruments. The issues that have been discussed in this chapter suggest that this search needs to be interrogated from a social perspective, asking the questions:

- which students are identified as high and low attainers by these instruments?
- what are the criteria by which students' responses are to be judged and to what extent are the students both aware of the criteria and in command of the resources needed to demonstrate that they have met them?
- is access to this awareness and command of resources differentially distributed among various social groups?
- how will use of the assessment instruments differentially affect the future of high and low attainers (considering issues of immediate curricular opportunities and access as well as qualifications for longer term education and employment)?
- what effects may implementation of these instruments have on teachers (considering issues of teachers' professionalism, control of the curriculum, and their immediate objectives as well as effects on the ways in which they interact with their students)?

Rather than attempting to find 'better' ways of assessing, a major aim of research into assessment in mathematics education that takes a social perspective must first be to understand how assessment works in mathematics classrooms and more broadly in education systems, and to understand what its consequences are for individuals and for groups within society. Understanding the ways in which assessment works and the consequences of assessment for both teachers and students must entail understanding the complex relationships between individual teachers and students and

the broader structures of the societies within which they act. It is only with such understanding of assessment as a social practice that the discriminatory effects of assessment practices can be recognised and challenged.

References

Anderson, J. (1989). Sex-related differences on objective tests among undergraduates. *Educational Studies in Mathematics, 20*, 165-177.

Baldino, R. R. (1998). School and surplus-value: contribution from a Third-World country. In P. Gates (Ed.), *Proceedings of the First International Mathematics Education and Society Conference* . Nottingham: Centre for the Study of Mathematics Education, Nottingham University, 74-81.

Bernstein, B. (1996). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. London: Taylor & Francis.

Björkqvist, O. (1997). Some psychological issues in the assessment of mathematical performance. In E. Pehkonen (Ed.), Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education (Vol. 1). Lahti: University of Helsinki, 3-17.

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.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to "realistic" mathematics contexts. *British Educational Research Journal, 20*(5), 551-564.

Bourdieu, P. (1990). *The Logic of Practice*. Oxford: Blackwell.

Broadfoot, P. M. (1996). *Education, Assessment and Society*. Buckingham: Open University Press.

Burkhardt, H. (1988). National testing - liability or asset. *Mathematics Teaching, 122*, 33-35.

Cockcroft, W. H. (1982). *Mathematics Counts*. London: HMSO.

Cooper, B. (1998). Assessing National Curriculum mathematics in England: Exploring children's interpretation of Key Stage 2 tests in clinical interviews. *Educational Studies in Mathematics, 35*(1), 19-49.

Cooper, B., & Dunne, M. (1998). Social class, gender, equity and National Curriculum tests in mathematics. In P. Gates (Ed.), *Proceedings of the First International Mathematics Education and Society Conference*. Nottingham: Centre for the Study of Mathematics Education, Nottingham University, 132-147.

Dowling, P. (1998). *The Sociology of Mathematics Education: Mathematical Myths/Pedagogic Texts*. London: Falmer.

Dunne, M. (1998). Pupil entry for Nathional Curriculum Mathematics tests: the public and private life of teacher assessment. In P. Gates (Ed.), *Proceedings of the First International Mathematics Education and Society*

Conference. Nottingham: Centre for the Study of Mathematics Education, Nottingham University, 148-157.

Ellerton, N. F., & Clements, M. A. K. (1998). Transforming the international mathematics education agenda. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics Education as a Research Domain: A Search for Identity – an ICMI Study* (Vol. 1). Dordrecht: Kluwer Academic Publishers, 153-175.

Galbraith, P. (1993). Paradigms, problems and assessment: some ideological implications. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education: An ICMI Study.* Dordrecht: Kluwer Academic Publishers, 73-86.

Garcia-Cruz, J. A., & Martinón, A. (1997). Actions and invariant schemata in linear generalising problems. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2). Lahti: University of Helsinki, 289-296.

Garcia-Cruz, J. A., & Martinón, A. (1998). Levels of generalisation in linear patterns. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2). Stellenbosch, 329-336.

Gill, P. (1993). Using the construct of 'levelness' in assessing open work in the National Curriculum. *British Journal of Curriculum and Assessment, 3*(3), 17-18.

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.

Hunter, M., & Monaghan, J. (1996). Some issues in assessing proceptual understanding. In L. Puig & A. Gutiérrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3). Valencia, 97-104.

Jaime, A., & Gutiérrez, A. (1994). A model of test design to assess the van Hiele levels. In J. P. da Ponte & J. F. Matos (Eds.), *Proceedings of the 18th International Conference for the Psychology of Mathematics Education* (Vol. 3). Lisbon, 41-48.

Karsenty, R., & Vinner, S. (1996). To have or not to have mathematical ability, and what is the question. In L. Puig & A. Gutiérrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3). Valencia, 177-184.

Kress, G. (1989). *Linguistic Processes in Sociocultural Practice*. (2nd ed.). Oxford: Oxford University Press.

Lawrie, C. (1998). An alternative assessment: The Gutierrez, Jaime and Fortuny technique. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3). Stellenbosch, 175-182.

Leder, G. (Ed.). (1992). *Assessment and Learning of Mathematics*. Victoria: Australian Council for Educational Research.

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*. Nottingham: Centre for the Study of Mathematics Education, Nottingham University, 26-33.

Lesh, R., & Lamon, S. J. (Eds.). (1992). *Assessment of Authentic Performance in School Mathematics*. Washington D.C.: American Association for the Advancement of Science.

Leung, S. S. (1994). On analyzing problem-posing processes: A study of prospective elementary teachers differing in mathematics knowledge. In J. P. da Ponte & J. F. Matos (Eds.), *Proceedings of the 18th International Conference for the Psychology of Mathematics Education* (Vol. 3). Lisbon, 168-175.

Love, E. (1981). Examinations at 16-plus. *Mathematics Teaching, 96*, 42-46.

Mathematical Association. (1990). *Cultural Diversity and Bias: Does GCSE Mathematics meet the criteria - A report on the 1990 papers*.

Morgan, C. (1996). Teacher as examiner: The case of mathematics coursework. *Assessment in Education, 3*(3), 353-375.

Morgan, C. (1998a). Assessment of mathematical behaviour: A social perspective. In P. Gates (Ed.), *Proceedings of the First International Mathematics Education and Society Conference*. Nottingham: Centre for the Study of Mathematics Education, Nottingham University, 277-283.

Morgan, C. (1998b). *Writing Mathematically: The Discourse of Investigation*. London: Falmer.

Morgan, C., & Watson, A. (forthcoming). Equity in teacher assessment of mathematics. .

NCTM. (1995). *Assessment Standards for School Mathematics*. Reston VA: National Council of Teachers of Mathematics.

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.

Noss, R. (1990). The National Curriculum and Mathematics: A case of divide and rule? In P. Dowling & R. Noss (Eds.), *Mathematics Versus the National Curriculum*. Basingstoke: Falmer, 13-32.

Orton, J., & Orton, A. (1996). Making sense of children's patterning. In L. Puig & A. Gutiérrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4). Valencia, 83-90.

Rickards, T., & Fisher, D. (1997). Assessment of teacher-student interpersonal behaviour: A seed for change. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4). Lahti: University of Helsinki, 56-63.

Ridgway, J., & Schoenfeld, A. (1994). *Balanced assessment: Designing assessment schemes to promote desirable change in mathematics education.* Paper presented at the EARI Email Conference on Assessment.

Romberg, T. A. (Ed.). (1995). *Reform in School Mathematics and Authentic Assessment*. New York: SUNY Press.

Ruthven, K. (1987). Ability stereotyping in mathematics. *Educational Studies in Mathematics, 18*, 243-253.

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.

Vinner, S. (1997). From intuition to inhibition – mathematics, education and other endangered species. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1). Lahti: University of Helsinki, 63-78.

Walden, R., & Walkerdine, V. (1985). *Girls and Mathematics: From Primary to Secondary Schooling*. Bedford Way Papers 24: Heinemann.

Watson, A. (1997). Coming to know pupils: A study of informal teacher assessment of mathematics. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4). Lahti: University of Helsinki, 270-277.

Watson, A. (1998). What makes a mathematical performance noteworthy in informal teacher assessment? In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4). Stellenbosch, 169-176.