ANNE WATSON

MATHEMATICS TEACHERS ACTING AS INFORMAL ASSESSORS: PRACTICES, PROBLEMS AND RECOMMENDATIONS

ABSTRACT. The purpose of this paper is initially to present findings which identify components of the practices of teachers acting as assessors of students' mathematics in the normal course of classroom work. At an informal level such practice is found to be complex and intimately related to every aspect of teaching and learning. It is found that even teachers who have undergone some assessment training may underestimate the role of interpretation of evidence, and questions about equity in the uses of teachers' judgements are raised in relation to awareness and practice. It is suggested that more care needs to be taken over the formation and use of 'professional judgements' within systems of assessment.

KEY WORDS: classroom decisions, mathematics assessment, mathematics teaching, social equity and mathematics, teachers' expectations

INTRODUCTION

It is common practice in mathematics in the UK to assign students to different curriculum tracks, a practice which usually commences in primary schools and develops into fully separated teaching groups by the age of 13 (Ofsted, 1994; Boaler, 1997). Tracking and setting decisions are made on a variety of grounds: national test results, teacher-written test results, other forms of teacher assessment such as judgements about motivation, behaviour, potential, mathematical ability and so on. Some of these judgements may be made without the need to justify them outside the classroom or the school. Teachers' monitoring, assessing, record-keeping and reporting of students are incorporated into the national statutory assessment requirements (HMSO, 1988). Their judgements, however founded, are used for official reporting to parents, deciding on examination entry, selection of teaching groups, ascribing levels to students throughout schooling and can also contribute to 20% of the final mathematics grade given at 16-plus. These changes have led to increased awareness about the methods and mechanisms of formal and summative assessment against given descriptors of mathematical performance (Askew et al., 1993; Brown, 1993).

Informal judgements and decisions which teachers make daily in their classrooms are also assessments (Dwyer, 1990) and can, through influencing expectations and interactions, significantly affect students' mathem-

Educational Studies in Mathematics **41:** 69–91, 2000.

atical experience (Ruthven, 1987, 1994; Lerman, 1989; Lorenz, 1982). It was felt to be important to find out more about the day-to-day mechanisms of judgement which operate in mathematics classrooms; that is, informal assessment done primarily for formative purposes. Teacher's views of, and decisions about, students' achievement are generally extended into statements of potential which can lead to different educational treatments. That is, informal assessment decisions can influence important judgements.

The purpose of this paper is to identify components of the practices of teachers, as assessors of students' mathematics in the normal course of their work, and to suggest directions for critique of such practices. This arises from a relatively small sample, and relates somewhat to the specific circumstances of UK teachers in that they have been trained in assessment procedures and have a statutory role as assessors. It is also possible that UK teachers, being accustomed to making discriminatory judgements about students' mathematics for tracking purposes, *expect* students to be different rather than similar. The role of mathematics as a social gatekeeper (Galbraith, 1993) makes it important to question the processes of assessment which are used to differentiate the curriculum, especially in view of the widely-held notion that objective statements of mathematical attainment are possible and can provide yardsticks against which to judge the reliability of informal assessments. Given increased use of continuous and contextual assessment of mathematics in many countries (Romberg, 1995; Van den Heuvel-Panhuizen, 1996), it is necessary to know how teachers see and undertake this role at an informal level. This aspect has been largely ignored in assessment theory and research. For example, in his development of a theory of assessment Webb (1993) mentions, but does not problematise, the need for interpretation of students' response in assessment situations. Rico et al. (1995) identify teachers' priorities in assessment, and their views of its purposes, but the difficulties they identify are due to examiners, students or procedures, not to the teachers' own practices. Some authors take teachers' biases into account in summative procedures (see later), but not in formative processes. This study, therefore, looks at teachers' practices of informal assessment and raises questions about how these might influence formative and summative judgements.

Teachers are expected to use their classroom experience to develop knowledge of their students' mathematical achievements and potential. They are therefore using classroom interactions (verbal and non-verbal) and interpretation as mechanisms for creating meanings. In this case the 'meanings' being created are the constructs of 'good' or 'weak' mathematics students. These constructs may involve judgements about motivation, effort, comprehension and potential, among other things, and are mediated by what the teacher values, how this is conveyed, how the student interprets and expresses what she thinks is valued, how the teacher responds and so on.

Various theoretical frames describe the classroom as an arena of such interaction: symbolic interactionism which focuses on the interpretation of symbolic communication as creating meaning (e.g. Blumer, 1969; Bauersfeld, 1988); social reproduction (e.g. Atkinson, 1985; Walkerdine, 1984) which emphasises how hegemonic values are expressed through schooling and determine classroom success; hermeneutics, which describe a cyclic process of interpretation and perception, each influencing each other (e.g. Brown, 1998); semiotics (e.g. Vile, 1997; Ernest, 1998) which focuses on the meaning and social exchange of signs; finally, Harre and Gillett's suggestion that conversation is the only appropriate unit of analysis of knowledge (Harre and Gillett, 1994; Ernest, 1998). There are others too, but apart from Morgan's work on teachers' interpretation of student-written text (1998) the application of social interaction and discourse theories to mathematics teaching and learning has largely been to illuminate how students learn, or fail to learn, mathematics rather than how teachers learn about students.

Teachers as assessors are required to produce statements about students' states of knowledge in the UK system, that is, to take a positivist view of students' knowledge. These statements are expressed as levels of attainment and predicted grades. As a framework, symbolic interactionism, with its emphasis on types of interaction and the assumption that a meaning is created through interactive processes, relates more closely to this requirement than more recent frames in which fluidity and ephemerality of meaning are of central importance. Interaction is taken to be communication of all kinds between student and teacher. All such frameworks could be used to question the validity of static statements of students' knowledge, but symbolic interactionism allows analysis of the relationship between process and statement. The research reported here was undertaken to find out more about these processes. In addition, symbolic interactionism has a significant history of use analysing the social complexities of classrooms (Delamont, 1976; Hargreaves, 1972; Voigt, 1994).

As an initial investigation into the field it was decided to take an approach similar to the discovery of grounded theory described by Glaser and Strauss (1967). If data collection continues alongside interpretation and theory-creation it is theoretically possible to continue collection until nothing new is being observed or heard. Constant comparison of results leads the researcher to search deliberately for contrasting or contradictory data, and emerging ideas, concepts and patterns can be tested out in the field through the continuing collection, interpretation and analysis. Thus it is an ideal approach where the aim is, as here, to *describe* the range of possible components of practice. Here the area being studied is the interpretative behaviour of the teacher relating to symbolic communication about mathematical learning.

The study of teachers' informal assessment practices

The research had three main parts: firstly the identification of practices of mathematics teachers acting as informal assessors; secondly, a critical study of how two teachers developed their views of some of their students during their first term with them (Watson, 1997); thirdly, a brief enquiry into peer-examination of professional judgement in school-based moderation practices (Watson, 1998a). This paper reports on the results of the first study.

Thirty teachers were selected from primary, middle and secondary schools, all of whom taught, or had in the recent past taught, 10, 11 and 12-year-olds. This level was chosen as it straddles transfer from primary to secondary school and hence includes special attention to the passing on of formative information, as well as summative statements. During the study statutory national testing was being introduced at age 11, and teachers were expected to make assessment decisions alongside this process. All teachers involved were therefore in some state of transition between using personal systems, systems common in their school, systems common within groups of schools and national systems. In relation to this all had received some training in assessment, usually in the form of guided assessment exercises, agreement trialling, and information about different methods and procedures.

The researcher, as an experienced teacher, typically spent a day supporting in the teacher's classroom establishing a relationship and some common ground. At the end of the day, or subsequently, a semi-structured interview of about an hour would take place about assessment practices, and was tape-recorded. It is arguable that interviews only represent teachers' views of practice and not their actual practice, however, since neither the students' mathematical understanding nor the teachers' thought processes are observable, interviews were used as the only source of data about the formation of judgements at this stage. The interviews were based around these two main questions:

How do you find out and recognise what children know and can do in mathematics?
Tell me about a particular student's mathematics.

Common experiences of students and incidents during the day were used to extend the discussion, as well as further probing and prompting of answers and information. Probes and prompts were informed by the analysis of previous interviews and were designed to elicit more information teachers' interpretations of their interactions with students, whether verbal, non-verbal or written. In particular similarities or differences between teachers' practices were probed, in accordance with the methods described by Glaser and Strauss (1967).

Love (1994), lists a variety of influences present in teachers' narratives: context, intentions, views of what is normal/abnormal, generalisability of illustrative incidents, what is left out, sense of power, time and place, interpretation of questions and answers, sense of coherence, assumptions about shared understandings, length and interconnectedness of the interview. All of these, he suggests, influence the meanings the interviewee intends to convey and the meaning the interviewer 'reads' into the interview, the interview itself being a field of social interaction in which interpretation and creation of meaning are taking place. Working with transcripts, without other data and without further discussion, it is a further interpretative task to take these largely unknown factors of unknown influence into account. The aim here is to describe *possible* practice, so the truth of what is said, and its real relationship with actual classroom events, are not issues; teachers can only talk of what they feel they know, and of possibilities of which they are aware, so their talk about assessment will be at least about possibilities, even if it is not about their own practice. Nevertheless, most teachers illustrated their responses with examples from their classrooms, probing was often about what happened rather than why it happened, and the second question was about named students. Hence there were few solely theoretical or hypothetical answers.

There is an interesting parallel here between the researcher's practice and the teacher's practice. Firstly the teacher is trying to find out what the student knows and can do in mathematics:

The teacher attempts to understand what the student knows through observation, speech and written work and to form these observations, through interpretation, into a description, using the classroom as the interactive arena. Theories and deductions are developed about what the student knows. The student describes and represents her knowledge in the written and other work. The student partly learns mathematics through an interplay of theory and deduction in various experiences, and the teacher comes to her conclusion about the student through a similar process. As well as the conjecturing and testing which goes on in learning mathematics, the teacher conjectures about the student through an interplay of theory, observation and inference in various situations.

Secondly the researcher is finding out what the teacher knows and does when assessing the student:

The researcher tries to understand the teacher's actions and intentions through speech (using a little observation) and to form these, through interpretation, into a description, using the interview as the interactive arena. Theories and deductions are developed about what teachers do. The teacher describes and illustrates her actions in her utterances. There is an interplay of conjecture and reality about the links and themes and issues identified, the way different kinds of action and phenomena are clustered and other aspects of theory-making. Teachers interpret the purpose of the research, which affects what they say, and they also have different foci of action which also affect what they say. The researcher conjectures about the teacher through these unavoidable layers of interpretation, which thus become a positive aspect of the work by being incorporated into theory-building.

The parallels would be complete if the teacher were aiming to find out what, in general, her students know and what range of ideas and perceptions it is possible for them to have. However, the teacher-as-assessor is expecting to construct an *actual state of knowledge* of individuals while the researcher constructing a *view of the possible practices* of a group. Another difference is that production of a description of possible practice (seen as a combination of people, context, actions, interactions, intentions and interpretations) is feasible, whereas a definitive statement of what someone else knows in mathematics is probably not. A further difference is that the researcher has to validate her findings explicitly and justify her interpretations to a critical audience, where the teacher may not have to justify or validate her views with others.

Interviews

The first interview question allowed teachers to talk about their systems and paperwork. This approach allowed related questions such as 'how did you decide that this student can do division with remainders?' to be asked. The question gave indirect insight into methods and influences, and was designed to lead into a conversation about types of evidence and judgements. A common supplementary was 'how do you know this?' It was not always used as bluntly as it is written here, more often arising naturally as a continuation of general discussion about the research.

The second question was designed to help teachers to focus in more detail on their methods. When using students as examples interviewees typically chose students who troubled them, either because they were particularly weak or strong in mathematics. Instead, the name of someone who had been seen in class, but who had not appeared to be extreme in any way, was offered. Teachers were encouraged to search their recollections of the student to see what kinds of incident sprang to mind and what they understood from it. (In this paper passages which are not referenced are quotations from interview transcripts.) For example, in the case of TC below such probing led to the observation that he collected information about students' interactions and working behaviour, and that he did this in his head, not necessarily on paper:

Q: . . . where does your information come from?
TC: It's my own judgement, my own thoughts about him and from watching him work.
Q: When you say 'watching him work' you don't stand there watching, do you?
TC: I suppose what happens is that you record things in your mind continually. What is it I record? I record how much work he is able to do, his interest in it, his attitude, how much supervision he needs, or encouragement, in particular do I have to speak to him to stay on task, and then I would also record how he is prepared to work without my help, or asking, who he works with, how well he works, whether he likes to work on his own or not, whether he asks questions that are not prompted by me . . . or being stuck.

Sometimes the researcher would try to shift the teacher from general statements about reactions to more specific statements about interactions and interpretations. For example, shifting LL's perspective from generalities to particular students evoked more detail about the kinds of interaction she initiated:

LL: You see these little faces working away and you ask them about things and watch the faces. And gradually you think 'there's no point me going on with this'.
Q: But what about someone like J, whose interpersonal skills seem to be bizarre and it is not so easy as reading, say, S's face, or B's face, from a different cultural background?
LL: But even with a child like J what you do is you just confirm while he is working, don't you? 'Are you alright? Is this OK?' and you just draw them in so you have to get the verbal response from them, don't you? And you have to spend more time visiting his work.

In this way teachers' recollections of interactions and how they interpreted them were explored, along with some of their underlying beliefs and motivations.

Identification of components of practice

The transcription stage was the first part of a cyclic interpretative process which was concurrent with collection of further data. The process continued with identification and coding of transcript contents, extension of the coding frame as more interview data was collected, the noticing of similarities and differences in the data, the clustering of features into sense-making groups, reclustering and frequently returning to old and new data for fresh insights. Additionally, interviewees' responses to features which appeared in earlier interviews were sought in order to develop a rich picture of similarities and differences. These analytical processes were complex and repetitive, involving many readings of the data and several stages of categorising, clustering and recategorising (Glaser and Strauss, 1967; Mellin-Olsen, 1993). A detailed description is not appropriate in a paper of this length, but it is important to mention that analysis did not end with the initial categorisation. Transcripts were also re-read at several later stages of the research.

Since teachers approached the task of answering questions in very different ways, ascribing codes required interpretation in order to say that different utterances were about similar things. Codes were ascribed liberally, a new code being created for every new feature which appeared in the data. Within codes there might be different views expressed, for instance some teachers were very hostile to the current assessment regime but others gave it partial support. Codes were assigned to components of assessment practice, in this case 'attitude to current system' (code 10.5 in Table I), rather than to particular viewpoints about components. In all, 66 features of practice were coded after several readings, and these were found to span all the data. These were then clustered together in related groups in order to name the main aspects and discuss them further. This clustering had to be done from a particular perspective. Several different frames were used to illuminate different aspects of the data (Watson, 1998a). The one presented here, in Table I, is based on a chronology of the practice of teachers-as-assessors. The clusters of components are arranged in the order in which they occur in relation to a normal sequence of lessons, according to the teachers' reports. Each teacher mentioned aspects of all the clusters apart from the last.

Teachers' views of mathematics and its teaching and learning were placed first, because they contribute to and predate how mathematics is taught, contribute to definitions of achievement, and inform criteria of attainment (Thompson, 1992):

SG: My own view about what a child who is good at maths should do þ well everyone has a model in their mind.

Although all formal assessment is carried out against a hierarchical typography of school mathematics contained in a national curriculum, some teachers talked of their own views of mathematics, their understanding of the structures of mathematics and their own experiences of learning mathematics as being different from this. For example, about half the teachers believed that it was possible for students to be able to do some mathematics one day and not be able to do it tomorrow:

SN: . . . whether they can do it next week is a different matter!

TABLE I

Components of practice of teachers-as-assessors

1. Views of mathematics
1.1 explicit hierarchy
1.2 implicit hierarchy
1.3 desirable order of teaching
1.4 child’s view of maths in general
1.5 teacher’s view of maths, teaching intentions
1.6 use of maths in another form
1.7 ‘can do’ today but not tomorrow
1.8 child’s interpretation of a maths task
1.9 speed and accuracy
1.10 misunderstandings
1.11 role of right or wrong answers
1.12 role of investigations

2. Systems
2.1 record format from other teachers
2.2 record accuracy from other teachers
2.3 usefulness of records from others
2.4 list of national curriculum content
2.5 list of what is covered in school
2.6 description of school system
2.7 self assessment
2.8 moderation
2.9 use of specific tasks for assessment

3. Oral
3.1 importance of articulation, whoever with
3.2 revealing remarks from children
3.3 communicating to child, individually or group
3.4 ‘telling back’ in own words

4. Written
4.1 do/write gap
4.2 homework
4.3 exercises
4.4 other types of written work

5. Tests
5.1 role of testing
5.2 when used
5.3 how results are used
5.4 problems of tests

6. Observation of students
6.1 using maths in context
6.2 seeing maths being ‘done’, watching work
6.3 body language
6.4 facial expression
6.5 observation of whole group
6.6 seeing maths in practical action, apparatus
6.7 learning behaviour/personality/social skills

7. Interpersonal/non-verbal knowledge
7.1 knowing the child’s traits, emotions etc.
7.2 interpreting signs
7.3 intuition, insight
7.4 ‘gut reaction’, instant judgement
7.5 having a mental image of particular child
7.6 information gained sitting next to the child

8. Emotions
8.1 confidence of child
8.2 fun/enjoyment
8.3 confidence that teacher has in a child
8.4 confidence of teacher with maths
8.5 teacher’s feelings about a particular child

9. Psychological attributes of student
9.1 child’s memory
9.2 ability
9.3 concept acquisition

10. Teacher’s attitudes and attributes
10.1 experience of children in general
10.2 experience of particular child’s mathematics
10.3 idea of what assessment should be
10.4 interpretation
10.5 attitudes to current system
10.6 memories of incidents with a particular child
10.7 teaching style

11. Making different kinds of judgement
11.1 subjectivity
11.2 fairness
11.3 about potential
11.4 peripheral, not seen as important
11.5 broad, approximate judgements
11.6 professional judgement

Many talked about the order in which they believed mathematics ought to be taught, or that they did or did not believe that it should be learnt in a particular order:

RD: Maths doesn't happen in an order that everybody does.

Hence cluster 1 in Table I was seen to underpin mathematical activity in their classrooms.

Systems and methods of assessment come next as cluster 2, because they describe the context within which the teacher-as-assessor works, followed by types of evidence and raw material which are expected to be generated by lessons:

SD: When we're doing something on the National Curriculum one way of doing it is to set them ten questions and then we give them a mark.
Q: What does that tell you?
SD: Not a lot but I suppose the idea is that if you've had a test you should have a mark.

The next four clusters describe the main kinds of observable evidence used for assessment purposes which accord closely to the literature on methods of assessment. Oral and written work and tests, (clusters 3, 4 and 5), can be planned in advance to a certain extent:

BM: I'm keen for them to explain it to me, what they are doing and why.
Q: What about those who just write the answer down?
BM: I do ask them to show their working out and see where the answer is from, but to take 97 from 101 in my head I would take away the 90 and then the 7. Written down I would want it done differently.

Observations of mathematics in context, unplanned demonstrations of mathematics, or of features of behaviour less directly related to mathematics, were put in cluster 6 as they are unexpected and may only be apparent on later reflection.

Some teachers had complete freedom in deciding how they would assess their students, some assigning levels of achievement by using 'what is in my head'; others worked within a system agreed by the school, using banks of special tasks or tests which were used in parallel groups; others believed that the only 'proper' way to do the statutory part of their assessment duties was to set a test or task rather like the national, externally-marked kind. In their informal practices some had detailed recording methods showing 'coverage', 'performance' and 'understanding' where others worked in a much less systematic way; some believed that they were assessing all the time, others spoke as if assessment was something they did deliberately at particular times:

KM: I have a system in which there are three levels of mark: one for 'has met', one for 'has done sometimes' and one for 'mastery'.

TC: I suppose I am building up a picture in my head all the time from what I see, what they say and what they can do . . . just from watching really.

Within these ranges of belief and practice there was a wide variety of suggestions of the kinds of classroom incident, planned or fortuitous, during which the teacher could find out something about student achievement, summarised in Appendix 1.

Interpersonal, emotional and psychological knowledge were put next in clusters 7,8 and 9 as they are more likely to be articulated *post facto* and taken into on reflection. Cluster 7 describes a group of aspects which are treated like evidence, in that the teacher uses them to make decisions, but they are not observable by others; this cluster moves the reader's attention from the observable to the psychological, in that it is about interpersonal interactions and judgements (Watson, 1995, 1997):

TC: I don't worry about what N can recall because that isn't a problem for him, what he does he remembers.
RS: It's a gut reaction really, just from knowing the child and the situation.

Clusters 8 and 9 describe further psychological aspects: the role of emotion in the mathematics classroom, and the cognitive and developmental psychological states which teachers talked of as underlying achievement and learning:

KB: Her parents are concerned about her, day dreamer. She sucks her thumb. She's happier and faster when she doesn't have to write things down.

Cluster 10 is a compendium of various other teacher-centred attributes which accumulate with time, such as memories and experience, which were mentioned as being brought to bear on judgements:

KM: I have a system which looks fairly mechanical but there are elements which are just picked up day by day as you interact with students.
AB: I know what I expect from children at this school.

It was noticeable that interpretation of students' work was raised only by three or four teachers as a very minor issue, whereas most of the relevant theoretical positions would place interpretation centrally in the functioning of human interaction:

GJ: If I don't know what they mean I ask them.
HJ: You can't always tell from what they write down, you have to ask them.

Finally, for cluster 11, a few teachers spoke reflectively about fairness, objectivity and other moral and ethical aspects of judgement; this seemed to be an evaluative discourse:

VI: To be fair you should set them all the same criteria, so you would have to set them all a test.
CB: I am pretty biased against tests because it doesn't give a true indication. I should include what I have seen of them in class, but I consider that I may have been wrong.

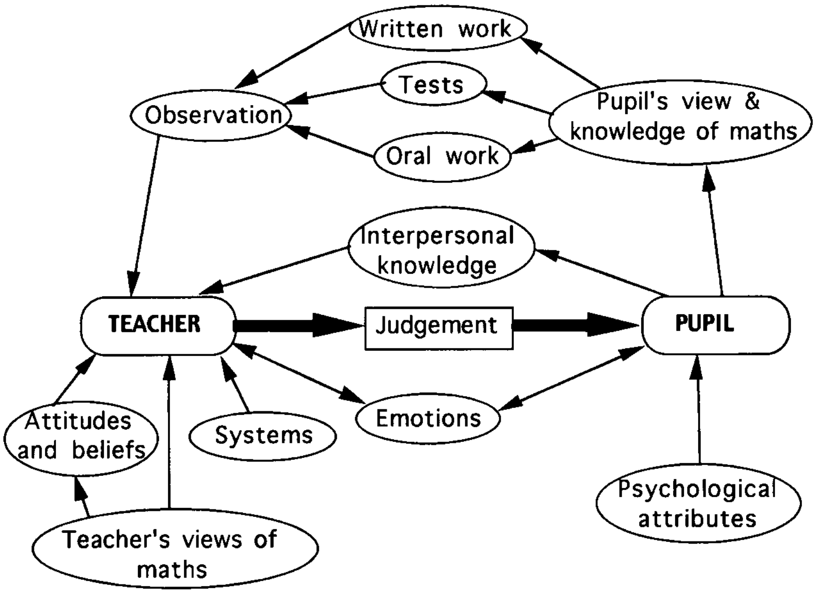

Figure 1. Teachers assessing students: interconnecting clusters from Table I.

Roughly, therefore, this clustering is chronological for the teacher who has existing views and a given situation and demands; then plans and acts; then notices, reads, hears and observes; then thinks about the observations, makes decisions bringing other things into account and reflects on them. As well as listing components of informal assessment practice the clusters provide categories within which questions about differences in practice might be raised, that is, a framework for comparison of practices.

Table I therefore forms an outline description of informal assessment practices according to the teachers in the sample.

Relating the components

Teachers' practice is more complex than the ordering of Table I implies. Figure 1 shows how various aspects of the role link together, combining components of practice in a diagram which reflects teachers' comments about the symbolic information they perceive in the classroom and the underlying influences on how they interpret it.

KM: You get an overall impression of the ones who are bright, you subconsciously file things away about a student. You might miss some, there are some who shine in different contexts; some did better in a maths challenge than they did in a test, and exams sometimes throw up odd results. It might be about ways of working, I have learnt to accept other ways of working than the one I was familiar with, for instance we now have to value divergent

thinking as well as convergent, but I do expect them to use my methods if they don't have any of their own which work.
LW: You give them the opportunities and the encouragement . . . based on years of doing this . . . you have to know how a child is going to react . . . it depends on the child, the mood, the circumstances . . . the child brings its whole being into the classroom . . . then I have a big bag of information about each child. Then you have to assess by looking at the work; you are assessing for a level, not for mathematical learning.

The diagram in Figure 1 evolved as a schematic means to record the interplay between the different clusters of assessment practices as they operate informally in the classroom. The arrows arise from indications of influence and causality found in the interview transcripts, and also from consideration of the theoretical backgrounds described at the start of the paper.

Since all teachers mentioned aspects of each cluster, this diagram is a way of seeing how components of teachers' informal assessment practices impact on teachers' judgements. Thus the model not only describes components of practice and their relationship but also provides a framework for future questions about when, where and how teachers interpret the mathematical work of their students.

The central actor in this model is the teacher, using observations of many kinds and interpersonal knowledge to make judgements about the student. Also contributing to the judgements are the teacher's views of mathematics, attitudes to a variety of aspects of her work, including systems, and an emotional element. The observations are of the actions of the student, influenced by the student's existing knowledge and views of mathematics. The generation of observable actions is partly influenced by psychological attributes, about which the teacher has made some inferences. In addition there are other aspects which contribute to the teacher's picture of the student, such as the social signs and skills used to convey interpersonal knowledge, and some emotional elements. This figure does not allow for possible overlaps between clusters, nor for those frequent classroom incidents in which emotional, psychological, mathematical, symbolic and attitudinal matters amalgamate.

Figure 1 is reminiscent of that produced by Koehler and Grouws (1992) to describe interactions in the mathematics classroom. It could be considered a fairly trivial grouping of what teachers said; however, its similarity to Koehler and Grouws' not only validates the model but also confirms that informal assessment is not an additional activity for teachers but is intimately connected with all interactive aspects of their teaching, even if they do not acknowledge this.

The immense amount of detail given in Table I about the components of Figure 1 ensures that it is non-trivial as a source of information.

It also provides a framework for connecting aspects of teachers' assessment practices. For example, critical analysis of the raw material used for judgements, namely oral interaction, written work, tests and observations, links to consideration of the influences *on* such judgements, in terms of systems, beliefs and attitudes, and the influences *of* the judgements on the students and hence to the considerations of fairness and equity. The details of influences on students are not considered in this study and would be an appropriate focus for further research.

Figure 1 also places the teacher's judgement about the student in the centre of the picture, so that the teacher's interpretative role is clearly central to assessment. Phrases such as 'gut reaction' (used to describe unconsidered judgements) and 'professional judgement' (used to describe a wide range of judgements made by teachers), which appeared frequently in the transcripts, may sidestep the implications of this role. This is not surprising, however, in view of the words of Blease (1983):

> Teachers are lead to believe, through their professional training, that they are able to make accurate judgements about children. They occupy a status in society whereby it is thought legitimate for them to make such judgements. In the heat of the moment they may more readily accept their own first impressions or the judgements of others without question – myth becomes reality (p. 124).

and, more recently, Dearing (1994):

> A policy which trusts more to teachers' professional judgement . . . will, coupled with an acceptance that schools are accountable to parents and society for their stewardship, produce the results we need (p. 27).

In mathematics everything is necessarily communicated in a symbolic form which is to some extent abstracted from what it represents. The lay person's view may be that correctly-written mathematics is representative of correctly-known and correctly-understood mathematics. This view places mathematics at one end of a spectrum of interpretivity which extends from the obvious and universally agreed to the subjective and contentious. In terms of school assessment performance of physical tasks, such as wiring a plug or scoring a goal, might be at one end of such a spectrum and aesthetic tasks, such as reading a poem or painting an abstract, might be at the other. To place mathematics at the 'obvious' end is to underestimate the role of interpretation involved. Written or spoken mathematics gives only a murky view of mathematical thought.

As well as a tendency to underestimate the importance of interpretation there are many other problems revealed in teachers' assessment practices, all of them exacerbated, particularly in the UK, by a systematic relationship to high-stakes assessment. Teachers' judgements made to enable pedagogic decision-making can be used within the system as if they are moderated and validated. Some of the associated problems are examined in the rest of this paper, and were found by returning to the transcripts and looking at similarities and differences in teachers' viewpoints about particular components of assessment practice (Watson, 1998b).

Raw material of judgement

An important set of components arising from the research was a list of the raw material on which teachers based their judgements. Transcripts were re-read specifically to find the starting points for teachers' utterances about individual students' mathematics. In the past these were called 'evidence' (Watson, 1995). They are a mixture of actions, mediated actions and the mediating stances of the teachers. Taking the teacher's perspective, since it is the teacher's senses which perceive evidence and the teacher who makes judgements, students' actions perceived by the teacher merge with teachers' predispositions to contribute to an interpretation. The following identified list does not exactly correlate with the Table I, but contains clusters 1, 3, 4, 5, 6, 7 and aspects which appear in other clusters. Seven types of raw material for judgement occurring frequently throughout the data are:

- oral evidence through the teacher overhearing conversations or commentary by students or through formal and informal pedagogic dialogue;
- written evidence in the form of exercises, tests, rough work, notes or writing about mathematical exploration;
- actions observed by the teacher while watching the student do practical activities or other work;
- unprompted use of mathematics while working on other mathematics, or in another context;
- behaviour and body language, as seen by the teacher;
- knowledge of child as a learner or in other respects;
- views about mathematics held by the teacher.

The first four types listed are widely recognised as raw material (Conner, 1991) and all are subject to interpretation and management problems, in contrast to the straightforward way they are presented in official literature (e.g. SEAC, 1991) and by individual teachers. Problematising this list marks a shift from description to critique of practice.

Problems relating to the raw material of judgement

In this section, problems relating to the raw material of judgement are described. These are from two sources: literature and comments made by some teachers who problematised aspects of practice, not necessarily their own.

Oral evidence, though highly valued by all the teachers, was reported to be time-consuming to organise. Language difficulties, diffidence or fear might prevent some students from offering it. It is rare to overhear useful remarks in a busy classroom, although such remarks often give insight into a student's thinking before they are able to record what they think on paper. Oral evidence does not give hard evidence to support a teacher's judgements, so that over-reliance on oral evidence may leave the teacher vulnerable to criticism from other teachers or inspectors. Most teachers said that they use oral interactions to make sure a child understands or to find out what led to errors in written work; some teachers said that they would be unwilling to believe a student understood unless they had heard the student explain 'in their own words'. Conversation is being used here to generate teacher's knowledge of individuals.

However, teachers' reported reliance on oral work can be seen in the light of Bernstein's work (e.g. 1971) on how middle class students are at an advantage in school because the elaborated codes of language are what they might be used to at home, where working class students are expected to communicate at school in a way very unlike the restricted codes used at home. Walkerdine (1988), however, suggests that the classroom language codes established by teachers are highly stylised. Though contrasting, both these views prompt a closer look at language forms in mathematics classrooms, leading at least to the realisation that 'explain how you did something', a common requirement in teacher-student discourse, is a rare form of speech outside school in any social grouping. Hence reliance on students' ability to use this form of language successfully involves expecting a keen awareness of different discourses as well as mathematical ability. We could look to Vygotsky (1978) to support reliance on oral work by saying that language is both the act and expression of thought, and knowledge is created through interaction via speech, although his work relates to language as a component of shared social practice, not necessarily between learners and teachers. Oral work would therefore give teachers access to how their students are thinking mathematically as well as influencing and constructing such thoughts. Teachers in the sample were not, in general, saying this; they were asking students to use speech to report mathematics already 'done'.

Written work was regarded as a safe reliable form of evidence which can be held up to scrutiny. Most teachers commented that they wanted more than 'right answers' in order to be convinced that students understood the work; they wanted oral evidence, or written workings and explanations as well. However, there was also wide recognition that many students had considerable difficulty in recording in writing what they could do mentally or practically (the ‘do/write gap’ of the code 4.1). Assessment based on written work has to be seen in the light of research resulting from increasing use of coursework in public examinations. Many writers (e.g. MacNamara and Roper, 1992) have shown that students can be very selective in what they write down, so that written work represents a highly edited view of their mathematical thinking. Sometimes this is an attempt to produce curtailed, terse, classical mathematics, but it can also be due to a failure to appreciate what is important, a ‘right-answer-only’ ethos, or an inability to find ways to represent abstract or intuitive thought on paper. Furthermore, Morgan (1998) has shown that teachers can vary widely in their interpretation of what written work represents.

Observed *actions* provide no permanent record of achievement. Organising observation in a busy classroom is difficult but such observations can reveal that the student is using particular methods, such as counting instead of using number bonds. Observation of actions depends on the teachers’ notions of how mathematical activity *might* be observable. Sometimes this is clear, such as when one sees a student use a ruler correctly and read off a measurement accurately. Other times it has to be interpreted, such as when a student is trying to make a cube from six squares and may appear to us to be doing it in an obscure way, but nevertheless succeeds. Other times, there is little to interpret; the student who is gazing motionless at a problem may or may not be thinking about it, and the thought may or may not be productive. On the other hand, avid writing or discussion may not indicate anything useful is being done. How the teacher interprets the actions can be influenced by many factors. In the examples above interpretation depends on what the teacher expects to see relevant to the mathematics, what the teacher expects from the particular student and what the teacher expects from students in general. It also depends on what is noted by the teacher. What is noticed is affected by preconceived impressions; and how teachers interpret what they see depends on the existing impressions of students’ ‘ability’. Casual observations are discussed further in Watson (1997).

Unprompted use of mathematics was highly regarded as a form of evidence, but difficult for teachers to plan for and only of use on rare occasions. There were different views about how long after being formally taught a topic one could regard its use as evidence of ‘knowing’; it was generally implied that there had to be some sort of time gap to be sure that the student had internalised the concept or method and was not relying on short-term memory alone; estimates varied from two weeks to six months as appropriate gaps. In a subsequent case studies in two classrooms *no* instances of unprompted use were noted by the researcher or the teachers (Watson, 1997).

Interpretation and situativity

The last three items in the list given above are not necessarily recognised by teachers as constituting raw material for judgement, but influence decisions through interpretation of the teacher's perceptions and the influence on perception of previous interpretations. The interpretation of signs and symbols expressed consciously or unconsciously by students, the ability of the student to produce acceptable texts, and the hermeneutic cyclic development of the teacher's knowledge of the student are all features of these elements of informal assessment. Theories of social interaction all alert us to the fact that these aspects of judgement are either subjective, with different perspectives leading to different interpretations, or determined by the norms and practices of the classroom in ways which may have little to do with the conventional subject matter of mathematics. Also affecting teacher's judgements are their interactions outside the classroom with colleagues and with authorities. External authoritarian structures (national curricula, examination syllabi, assessment criteria, school inspection systems) influence the development of mathematical meaning and knowledge within the classroom; they also influence the teacher's perspective of what is required and what is acceptable in school mathematics.

While teachers generally gave little importance in their interviews to the problems of interpretation, they *did* imply a recognition that students' knowledge of mathematics was, to a certain extent, situated. This they saw as making it difficult to give summative statements, which they are sometimes required to do, about what they 'know and can do'. The awareness of situativity was expressed in phrases such as:

RD: What they can do today they might not be able to do tomorrow
JG: I can't be sure they know until they have used it in some context
KP: Sometimes they can do it before they can write it
SN: They can do things with me that they can't do with the next teacher

The sense of situation is fairly naive, limited to classroom, time, teacher and teaching context, but the underlying implication is that knowledge is not fixed and accountable. About half of the teachers, while confident about their own judgements, expressed frustration at having to produce summative statements which could only be, at best, flawed snapshots of the dynamics of learning mathematics:

IW: I don't think you can formalise summative assessment . . . you can provide different assessment activities, that has value in a limited way.
LL: I object to giving a level just for the sake of it. I don't think there is any point in testing

unless it is diagnostic. If you are talking about standard tests I think they could be a lot better. The levels are too far apart and they are only a snapshot.

However, it was frequently implied that if only one had more time to spend with each student, such a true summative picture might be possible.

CONCLUSION: TEACHERS' ASSESSMENTS AND EQUITY

Teacher's assessment practices, both informal and formal, are threaded through with problems of observation, perspective, interpretation and expectation. Hence it seems entirely possible for students to be assessed differently by different teachers with different views of mathematics, of the statutory requirements, of what makes a 'good' student, of interpretations of the student's work, and so on. Recalling that the teachers in this sample were trained in assessment and were contributing to high-stakes decisions about students' mathematics, this analysis shows clearly features of assessment which must be examined and challenged to ensure equity. There are also clear research requirements to study the effect of teachers' informal assessments on students' learning behaviour and mathematical attainment, and their relationship to high-stakes assessments.

Several writers on assessment recognise this as an issue. The usual suggestion is that inequity can be reduced by careful interpretation of criteria and some form of triangulation, such as use of a range of evidence for assessment, institutional use of exemplary portfolios and discussion of judgements with colleagues (Clarke, 1996; Gipps and Murphy, 1994; Ridgway and Schoenfeld, 1994; SEAC, 1991; NCTM, 1989). These measures may be applied to formal summative uses of teacher assessments but cannot be applied explicitly and systematically to teachers' informal judgements. Use of a range of evidence and discussion could, however, be encouraged through systems and management. Further, it would be helpful to encourage suspension of judgement, avoidance of discriminatory pedagogic decision-making, and deliberate search for alternative interpretations of evidence. In other words, knowledge of the role of interpretation, an understanding of the potential for inequity, and personal doubt, should be components of professional decision-making.

Another approach might be to 'check' all judgements by devising closely-focused test situations which could confirm or contradict teacher judgements, differentiating test style between mathematical situations in which strong symbolic conventions and right answers apply and those which allow a more individualistic approach (Heuvel-Panhuizen, 1996). However, this relies on sophisticated test-devising skills and the belief that mathematical knowledge is ultimately measurable and amenable to summative statements.

A more realistic approach might be to accept that the best a teacher can do is to behave as if her interpretation of students' responses gives her adequate but tentative, ephemeral information for teaching purposes, retaining an open mind and avoiding irrevocable decisions such as tracking, stereotyping and labelling.

ACKNOWLEDGEMENTS

The comments of the anonymous reviewers have helped the presentation of this work immeasurably.

APPENDIX 1: TYPES OF ASSESSMENT OPPORTUNITY USED BY SAMPLE TEACHERS

Observed use of mathematics
Mathematics used as expected in a closed question
Mathematics used in adapted form, or as part of doing other work
Mathematics used while doing practical or investigational work
Simpler mathematics used while doing more complex mathematics

Explanations
Explaining to the teacher in the students' own words
Explaining to another student

Responses
Verbal response to teacher-led questioning
Verbal response to open prompts, e g. 'tell me about . . ..'
Written or oral response to similar, simpler, slightly different or harder examples, or examples where questions are asked in other ways
Written or oral response to questions which reverse the order, e.g. 'what numbers multiply to give 48?" instead of '$6 \times 8 = ?$'

Expressions of insight
Student expressing insight while working on an intended area of mathematics
Insight while working on another area of mathematics
Insight while communicating student-to-teacher or student-to-student

Tests
Tests, teacher-written
Impromptu use of testing questions
Use of bank of test items
Testing as part of published scheme or programme of work
Tests written by students for their class;
Test times: pre-topic, during topic, post-topic, a few weeks later, six months later, fixed times during year, as preparation for national tests, when the student feels ready for a test

Self assessment
Self assessment: summary of achievement
Students making up their own examples to demonstrate understanding

Errors
Analysis or discussion of errors
Errors revealing partial understanding

Other activities
Activities which use mathematical knowledge or processes, or both, and are expressed through paper, observation, verbal, investigative or practical work, either as normal classroom work or as a special assessment task.

REFERENCES

Askew, M., Brown, M., Johnson, D., Millett, A., Prestage, S. and Walsh, A.: 1993, *Evaluation of the Implementation of National Curriculum Mathematics at Key Stages 1, 2 and 3,* School Curriculum and Assessment Authority, London.

Atkinson, P.: 1985, *Language, Structure and Reproduction,* Methuen, London.

Bauersfeld, H.: 1988, 'Interaction, Construction and Knowledge: alternative perspectives for mathematics education', in D.A. Grouws and T.J. Cooney (eds.), *Perspectives on Research on Effective Mathematics Teaching vol 1*. Erlbaum, New Jersey.

Bernstein, B.: 1971, *Class, Codes and Control, vol 1* RKP, London.

Blease, D.: 1983, 'Teacher Expectations and the Self-fulfilling Prophecy', *Educational Studies* 9, 123.

Blumer, H.: 1969, *Symbolic Interactionism: perspective and method,* Prentice-Hall, Englewood Cliffs.

Boaler, J.: 1997, *Experiencing School Mathematics: teaching styles, sex and setting*, Open University Press, Buckingham, UK.

Brown, M.L.: 1993, 'Assessment Philosophy and Practice in the UK', in M. Niss (ed.), *Cases of Assessment in Mathematics Education*, Kluwer, Dordrecht.

Brown, T.: 1996, 'The Phenomenology of the Mathematics Classroom', *Educational Studies in Mathematics* 31, 115–150.

Clarke, D.: 1996, 'Assessment' in A. Bishop (ed.), *International Handbook of Mathematics Education*, Kluwer, Dordrecht.

Conner, C.: 1991, *Assessment and Testing in Primary School*, Falmer, London.

Dearing, Sir R.: 1994, *The National Curriculum and its Assessment*, School Curriculum and Assessment Authority, London.

Delamont, S.: 1976, *Interaction in the Classroom*, Methuen, London.

Dwyer, C.A.: 1990, 'Trends in the Assessment of Teaching and Learning', in P. Broadfoot (ed.), *Changing Educational Assessment*, RKP, London.

Ernest, P.: 1998, *Social Constructivism as a Philosophy of Mathematics* SUNY, New York.

Galbraith, P. (1993): 'Paradigms, Problems and Assessment: some ideological implications', in M. Niss (ed.), *Investigations into Assessment in Mathematics Education*, Kluwer, Dordrecht.

Gipps, C. and Murphy, P.: 1994, *A Fair Test*, Open University Press, Buckingham.

Glaser, B.G. and Strauss, A.L.: 1967, *The Discovery of Grounded Theory: strategies for qualitative research*, Aldine, Chicago.

Hargreaves, D.: 1972, *Interpersonal Relations in Education*, RKP, London.

Harre, R. and Gillett, G.: 1994, *The Discursive Mind*, Sage, London.

HMSO, Her Majesty's Stationery Office: 1988, *The Education Reform Act*, HMSO, London.

Koehler, M. and Grouws, D.A.: 1992, 'Mathematics Teaching Practices and Their Effects', in D.A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning*, MacMillan, New York.

Lerman, S.: 1989, 'Investigations: where to now?' in P. Ernest (ed.), *Mathematics Teaching: the State of the Art*, Falmer, London.

Lorenz, J.H.: 1982, 'On Some Psychological Aspects of Mathematics Achievement Assessment and Classroom Interaction', *Educational Studies in Mathematics* 13, 1–19.

Love, E.: 1994, 'Mathematics Teachers' Accounts Seen as Narratives' in L. Bazzini (ed.), *Proceedings of Fifth International Conference on Systematic Cooperation Between Theory and Practice in Mathematics Education*, Grado, Italy, pp. 43–155.

MacNamara, A. and Roper, T.: 1992, 'Attainment Target 1 – is all the evidence there?', *Mathematics Teaching* 140, 26–27.

Mellin-Olsen, S.: 1993, 'Where is the Student as a Subject?', in M. Niss (ed.), *Investigations into Assessment in Mathematics Education*, Kluwer, Dordrecht.

Morgan, C.: 1996, *An Analysis of the Discourse of Written Reports of Investigative Work in GCSE Mathematics,* Unpublished PhD thesis, University of London, Institute of Education.

Morgan, C.: 1998, *Writing Mathematically: the discourse of investigation,* Falmer, London.

NCTM, National Council for Teachers of Mathematics: 1989, *Curriculum and Evaluation Standards for School Mathematics*, NCTM, Reston, VA.

Ofsted: 1994, *Science and Mathematics in Schools: a review*, HMSO, London.

Rico, L., Fernandez, F., Gil, F., Castro, E., del Olmo, A., Moreno, F. and Segovia, I.: 1995, 'Teachers' Conceptual Frameworks of Mathematics Assessment', in D. Carraher and L. Meira (eds.), *Proceedings of the Proceedings 19th Conference of the International Group for Psychology of Mathematics Education*, Recife, Brazil.

Ridgway, J. and Schoenfeld, A.: 1994, *Balanced Assessment: designing assessment schemes to promote desirable change in mathematics education*, EAR1 email conference on Assessment.

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

Ruthven, K.: 1987, 'Ability Stereotyping in Mathematics' *Educational Studies in Mathematics* 18, 243–253.

Ruthven, K.: 1994, 'Better Judgement: rethinking assessment in mathematics education', *Educational Studies in Mathematics* 27, 433–450.

SEAC: 1991, *NC Assessment: a moderator's handbook 1991/2, Key Stage 1*, HMSO, London.

TTA, Teacher Training Agency: 1997, *Standards for the Award of Qualified Teacher Status*, TTA, London.

Thompson, A.: 1992, 'Teachers' Beliefs and Conceptions: a synthesis of research', in D.A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning*, Macmillan, New York.

Van den Heuvel-Panhuizen, M.: 1996, 'How pencil-and-paper assessments can provide footholds for further instruction', paper presented to AERA symposium, New York.

Van den Heuvel-Panhuizen, M.: 1996, *Assessment and Realistic Mathematics Education*, Freudenthal Institute, Utrecht.

Vile, A.: 1997, 'What can Semiotics offer Mathematics Education', *Proceedings of British Society for Research into Learning Mathematics*, Oxford, 93–105.

Voigt, J.:1994, 'Negotiation of Mathematical Meaning and Learning Mathematics' *Educational Studies in Mathematics* 26, 275–298.

Vygotsky, L.S.: 1978, *Mind in Society: the development of higher psychological processes*, Harvard University Press, Cambridge, Mass.

Walkerdine, V.: 1984, 'Developmental Psychology and the Child-centred Pedagogy: the insertion of Piaget into early education', in J. Henriques (ed.), *Changing the Subject: Psychology, Social Regulation and Subjectivity*, Methuen, London.

Walkerdine, V.: 1988, *Mastery of Reason*, London, Routledge.

Watson, A.: 1995, 'Evidence for Pupils' Mathematical Achievements' *For the Learning of Mathematics* 15, 16–21.

Watson, A.: 1997, 'Coming to Know Pupils: a study of informal teacher assessment of mathematics' *Proceedings of the 21st Conference of the International Group for Psychology of Mathematics Education*, University of Helsinki, Finland.

Watson, A.: 1998a, *An investigation into how teachers make judgements about what pupils know and can do in mathematics*, Unpublished DPhil thesis, University of Oxford, UK.

Watson, A.: 1998b, 'Potential Sources of inequity in teachers' informal judgements about pupils' mathematics', in P. Gates (ed.), *Proceedings of the First International Mathematics Education and Society Conference*, University of Nottingham, UK.

Webb, N.L.: 1993, 'Visualising a Theory of Assessment' in M. Niss (ed.), *Investigations into Assessment in Mathematics Education*, Kluwer, Dordrecht.

Centre for Mathematics Education Research,
University of Oxford,
Department of Educational Studies,
15 Norham Gardens,
Oxford OXZ 6PY,
U.K.