

The Perceptions and Practice of Assessment in Mathematics Education of Teachers of Years 3 and Year 6 Students

Maxine Pfannkuch
The University of Auckland
<m.pfannkuch@auckland.ac.nz>

This is a report on part of a larger study about teacher assessment practice in New Zealand. The interviews with teachers of Year 3 and 6 students are analysed. These interviews indicate that teachers are using a variety of assessment strategies for a mastery-based system of assessment. The judgement of students' mathematical performance appears to be dominated by the belief that all students must feel that they are achieving. It is conjectured that a system effect determines teachers' assessment in the school environment.

Recent changes in mathematics curricula have clearly signalled that changes in teaching approaches and assessment approaches are desirable. The literature would suggest that much of the curriculum innovation has failed to take root in the classroom mainly because of system and teacher factors. The system factor dominates and determines classroom practice in assessment since the methods and purposes of assessment are deeply embedded in the school system of a country. Thus changes in assessment cannot be effected unless "old styles of thinking" (Romberg et al., 1990, p. 26) are divested in the institutional system. It is believed that this current lack of cohesion between the goals of the mathematics education curriculum and assessment philosophy and practice inherent in the system, has stifled changes in teaching. However "assessment involves a multitude of deep and difficult theoretical and practical problems" (Niss, 1993, p. 27) which will not be easy to resolve.

The teacher factor identifies teachers' perceptions, beliefs and conceptions of mathematics and assessment as playing a critical role in classroom practice (Watson, 1998; Kyriakides, 1998). According to Thomas (1998, p. 6) teachers' knowledge of mathematics and of their students strongly affects their teaching. "It shapes how they select activities and resources, how they present material in class, how they interact with students and how they assess student progress." Furthermore there is still a limited understanding of teachers' beliefs about assessment and the ways in which they practise and evaluate their assessment methods (Philippou & Christou, 1996; Warren & Nisbet, 1999). Bjorkqvist (1997) also believes that teachers' understanding of assessment in mathematics is shaped by teacher interaction with other teachers, by government policy and by society which evolve into a sociomathematical norm. Thus the system factor and teacher factor are complexly interconnected. Therefore any study on teachers' perceptions and practice of assessment in mathematics must take into account the school and system culture.

Background to Study

In 1992, for the first time in New Zealand, a mathematics curriculum (Ministry Of Education, 1992) was developed for the whole school age cohort. This curriculum was divided into six content strands and eight levels of achievement. Achievement objectives for each level and each strand were specified. In order to achieve these objectives suggested learning experiences and sample assessment tasks were included as exemplars for teachers. The introduction of this new curriculum was promoted by politicians as providing accountability through specified objectives, and by educators as a means of changing traditional mathematics teaching and assessment. These pronouncements caused members of the Mathematics Education Unit at The University of Auckland to wonder if this curriculum

or future curricula would produce the desired outcomes in the classroom. Furthermore when the Unit was asked to comment on whether the changes had improved the teaching of mathematics, we had no data on which to compare and evaluate the effects of the new curriculum. Thus it was proposed that a 'snapshot' of four areas – technology, the teacher, the student and assessment - would be taken at ten-yearly intervals at the Year 3, 6, 8, 10 and 13 levels (i.e. 7, 10, 12, 14 and 17 year old students). This paper reports on the first 'snapshot' of the assessment practice of teachers of Year 3 and 6 students as the new curriculum was being implemented in the classroom.

Method

Research (e.g., Philippou & Christou, 1996) indicated that what teachers reported did not always match with what they actually did in practice. A decision was made to interview teachers and gather documentation, as my aim was to partially understand the whole assessment process practised by a teacher. The protocol for the interview was based on the assessment framework developed by Niss (1993). This framework is a structural analysis of assessment in mathematics education and is grouped into the two categories of purposes and modes. According to Niss the purposes of mathematics assessment are "the provision of information, the establishing of bases for decisions and actions and the shaping of social reality" (Niss, 1993, p. 6). A given assessment mode includes the whole spectrum of components that model the assessment process. Thus these components, which include the subject, the objects, the items, the occasions, the procedures and circumstances, the judging, the recording, and the reporting, formed the basic structure for the interviews and the collected documentation. Questions were also asked about the rationale for changing assessment practice in mathematics.

Table 1
Subject and School Data

Subject Code	Year	Gender	Ethnic Group	Age Group	Years Teaching	School Status	School Size
3.1	3	F	NZE	20-29	2	M	M
3.2	3	F	NZE	50-59	15	L	M
3.3	3	F	NZE	50-59	20	H	M
6.1	6	F	NZE	30-39	12	M	M
6.2	6	F	NZE	40-49	16	L	M
6.3	6	F	NZE	50-59	30	L	M

Note. Ethnic Group: NZE=New Zealand European; School Socio-Economic Status: L=low, M=middle, H=high; School Size: S=small(<400), M=medium, L=large(>1200).

In order to obtain a cross-section of assessment practice from the schools in the greater Auckland area a random sample of three teachers at each level was selected. From November 1994 until September 1996 interviews were sought and conducted with three teachers at each of the designated levels. A one hour in-depth interview was audio-taped and samples of the teacher's assessment tasks, record book and report form were collected. In some cases the documentation was not made available even though some of it was seen by the interviewer. The interviews and documentation were analysed independently by me and another researcher with primary teacher experience. We subsequently arrived at a consensus about our interpretation of the interviews.

Since participation in this study was voluntary the teachers in the sample may not be representative of the teacher population. Demographic details about the subjects and their schools are listed in Table 1.

Results of Interviews

The results are based around the Niss (1993) framework in order to present the linkages and connections within the assessment system of a school. This is not a full report of the results.

The Report to Parents

All the school reports to parents had a mathematics achievement rating on a three-point scale which could be described as reflecting the normal curve by categorising students as below average, average, or above average. However, the schools used their own terms for these categories. For example, School 6.1 used S=strength, Comp=competent, ED=experiencing difficulty, while School 6.2 used A=excellent, B=satisfactory, C=having difficulty. Two schools (3.1, 6.1) gave mathematics achievement ratings in the three areas of number facts, applications and problem solving whereas the other schools gave one global rating.

Record Keeping

This three-point rating scale was reflected in the teachers' mark books with the predominant use of a tick, dot and a cross for the three ratings. Some marks were recorded for the assessment of number facts. Comments on individual student achievement and progress ranged from detailed comments in the mark book margins to no comments. The mark books, in all cases, had a list of curriculum objectives, broken down into 'manageable' parts for a mastery assessment procedure. These curriculum objectives were developed by either the school, the teacher, or a commercial publisher based on their interpretations of the mathematics curriculum objectives. On remarking that some of the curriculum objectives had been broken down to into smaller parts, 6.3 replied:

Oh yes they have got to be in and some of the teachers have been complaining because they want these [the objectives] to be broken down even more because some children don't show any progress.

This was a dominant theme that the objectives should be sufficiently broken down so that every student was able to demonstrate progress. It was not clear always to the interviewer how the teachers obtained a global rating from their mark books for the report. However 6.3, when pressed, agreed that she made a visual judgement of the number of ticks, dots and crosses in her mark books. Three teachers (3.3, 6.2, 6.3) also had a portfolio of mathematics work for each individual student.

The Assessment Task

The main forms of assessment were written, observation and interview. For the written assessment many gave the same pre- and post-test. The dominant assessment task was the pencil and paper test based on the objectives, with some practical tasks particularly for measurement, statistics projects and construction of an object. Most of the tasks were photocopied from or based on commercial resources. Very few were designed by the teacher. The actual assessment tasks concentrated on facts, procedures and applications. Very few tasks invoked higher order thinking or could be described as problem solving. Units of work and thus the assessment were based around one content strand. Many assessed tasks were an integral part of the teaching. Students would be given mathematical activities for a content

strand and if the teacher informally observed the ability of the student to achieve that task it would be noted, sometimes in their record book, sometimes in their head.

90% of it is mental notes, I mean it would be really nice if we could record everything but I've tried something with this jolly thing on my knee and a piece of note paper and trying to teach the kids at the same time ... (3.1)

Criteria for Judgement of Students

All the teachers operated in their classroom by putting students into three or four ability groups. The pre-test or the teacher's professional judgement determined which group a student was placed in for a unit of work, although in reality the nature of the groups did not change much throughout the year. None of the teachers used the students' previous class results or placed value on the other teacher's judgement for determination of their ability group structure as they believed students could change or forget their work in the six week holiday period. Comments such as the following show a strong belief in their ability to judge the appropriate group for a child.

I don't really need what the teachers from the years before said because I can tell so quickly where a child is at. (6.3)

No, to tell you the truth I never looked back to see what happened last year because they change so much. I just do it on the first couple of lessons ... I will put out a level 1 and a level 2 activity and just see which ones they cope with and I assess from there which group they will have to work in. (3.2)

The over-riding criteria for all teachers is what they called their 'professional judgement'. The teachers chose the commercial resource to assess their students. This resource could be modified to more suit their students or to enable their students to show progress.

In this pre-test for algebra – it's not a straight test from the topic book. It's been changed because we wanted to cut out as many words as we could. (6.3)

There was a general belief that the assessment task should be suitable for their students, that students must feel they are succeeding, not failing.

I couldn't give them all the same test at the end of the day, no. Because otherwise they've got that feeling of failure which we don't want them to get because they must enjoy maths and feel they have achieved something. (3.2)

Generally each ability group mastered an objective on a differentiated level of achievement since teachers admitted they had different expectations for the different groups. Thus two students in the same class who were said to have mastered an objective could not be considered to be able to do similar tasks. The criteria or benchmarks for achievement of objectives were professional judgement, against a background knowledge of textbook examples and in two cases BSM (Beginning School Mathematics) benchmarks. However the overriding criterion was that students must feel that they are progressing and hence the assessment appeared often to be against the perceived potential of the students. For instance, 3.1 realised during the interview that she taught her highest group (diamonds) differently to the lowest group (triangles). On questioned about differentiated and higher expectations in assessment she replied:

Yes I suppose I would, if I teach them differently, I'd assess [differently] ... definitely ... for a child in the diamonds I might say 'oh, they could do a little bit better than that' ... whereas with the triangles it might be 'oh, at last, and right let's forget it'.

Four teachers (3.1, 3.3, 6.1, 6.3) worked in syndicates (with other teachers) and either used standardised assessment tasks developed by the syndicate or used their own tasks. In both situations there was no discussion on whether their judgement of work was similar to another teacher's judgement. It seemed that teachers shared assessment tasks but did not share

outcomes. Two teachers (3.2, 6.2) worked alone. Thus achievement of objectives by students probably could not be used for comparison of students across rooms (the same year level) in a school.

Yes, I would compare a child with the children in my room but I haven't thought to compare across the rooms ... you very much stick to your own little [world] ... we [the syndicate] all know we are doing statistics for five weeks and if someone says 'hey, this is a really good activity try this one' then we share ideas like that but we very seldom share the outcomings of the children's work. (3.1)

In fact the teachers stated that the standards they expected from students would be quite different to standards expected by another teacher in the same school and definitely would be different from school to school.

We do talk to each other but I wouldn't guarantee that our standards are the same. I think, no they wouldn't be I am sure. There could be a big difference ... presentation and everything, yes. A big difference. (6.2)

They agreed that the school report could be differently interpreted by teachers since this was an aspect that was not discussed.

I've never discussed it with another teacher. We write our own reports ... they [school hierarchy] just said this is the report format. (3.1)

The Rationale for Changing Assessment Practice in Mathematics

All the teachers were able to rationalise the changes from a classroom practice perspective. For example:

Well, I think something needed to be done about the assessment of maths ... I was teaching it before and it was very airy fairy ... something was done and we've taken it on board and gone with it (3.3)

One teacher (6.2) said that some new assessment methods including how they analysed students' responses in traditional tests enabled her to learn more about students' learning. Some believed that the drive for change was emanating either from the school hierarchy, BOT (Board of Trustees), ERO (Education Review Office), parents, or politicians. They believed that there was more accountability sought now.

Since the new curriculum came in ... I have to show concrete facts ... to BOT and ERO whereas before I didn't have to do that. (3.2)

There was a general feeling that an external source such as the syndicate leader, the curriculum, or ERO dictated that there would be a change in assessment, told them how to do the assessment and they did what they were told. However there was an underlying belief that the changes might not improve the performance of the students.

I think it has been driven from above, from the ministry ... and we think 'oh we've got to do this' ... I think we are all afraid that it's going to be imposed from above because we feel we know what's best for our children. (3.2)

From that book [curriculum] I guess they just want to improve the children's standard of work ... but is that really important, do we really need to notice? ... we can't really do much about their brains. (6.2)

School policy stated that 3.3 had to assess using the curriculum objectives.

Because I've had to and they've been dictated to me ... Well, they're trying to make a uniform system where we are all doing the same thing.

None were able to articulate a mathematics education research or an international mathematics assessment perspective. Their understanding of what assessment was actually measuring or what they sought to measure in mathematics learning was at a practice level rather than a theoretical level.

Discussion

Two themes have emerged from this particular analysis. The first theme is the school effect, the second is the education of the teacher. Both these are system effects, not teacher effects.

The first theme of the school effect expresses itself in the requirement for reports, the way teachers operate in that school's system, the way teachers work in their classrooms and the value the school places on having an educated work force. The way teachers operate in their school system is based around a system of either working in a syndicate or working alone. Working in syndicates could provide opportunities for teachers to share assessment tasks, to discuss expected outcomes, and to share students' assessment output for discussion of judgement criteria. It appears teachers only share assessment tasks in this system. Also the decision makers in the school determine the form and scope of the report information. There would appear to be no discussion on what was being assessed, why it was being assessed and why it needed to be reported to parents.

The way these teachers work in their classrooms is dominated by the belief that assessment must show that each child has progressed or achieved success. Therefore each student is an individual and the achievement rating reflects the teacher's assessment, partially benchmarked against what she feels is the student's potential. Thus an achievement rating cannot be used for comparison of students within a class, between classes in a school or between schools.

The school must provide accountability to parents, ERO and the community about the standard of mathematics achievement of individual students and groups of students. Therefore teachers' assessment is focussed on producing evidence of student progress (e.g., the pre- and post-test, the mastery of objectives) and producing a global rating on a three-point scale in order to demonstrate student achievement in mathematics for the system. This achievement rating is largely based on an individual teacher's professional judgement of the tasks to be used and the interpretation of the students' responses. These criteria for judging student achievement should be challenged, particularly when professional development in assessment does not seem to include discussion and comparison of teachers' judgements on incidents or tasks that are assessed. According to Watson (1998) teachers need to be critically aware of their judgement processes and how such factors as their first impressions of the mathematical abilities of students and their view of mathematics may influence their perception of students' achievement in mathematics. Watson furthermore stated that, once teachers' beliefs about students had been established, the teachers reinforced these beliefs by only remembering favourable incidents, and in spite of mounting evidence to the contrary there was a reluctance to change original opinions of the students' abilities in mathematics. Therefore professional judgement and impartial assessment practice is a concern that perhaps should be addressed within the teaching syndicates as part of the ongoing professional development of teachers.

All the schools were up-to-date with the new curriculum in that they were using the curriculum objectives and using a variety of assessment tasks. However there was no questioning of the new assessment practice. There was a sense that teachers felt they were being dictated to by someone else, be it the school hierarchy, ERO, BOT, parents, politicians, the curriculum, or mathematics resource teachers. This sense of powerlessness was expressed as *show me what to do and I'll do it*. There was no sense of energy, debate, evaluation of, and judgement on the assessment practice advocated in the mathematics curriculum. In fact teachers felt overwhelmed by the administrative demands of objective-based assessment and the ongoing changes in all curricula areas. Providing assessment task resource material and showing teachers how to use it and mark it gives procedural knowledge or empirical experience. What appears to be lacking is the theoretical knowledge of why changes should

be made to assessment practice and, for example, of the difference in lower, middle and higher order thinking tasks in mathematics (de Lange, 1995).

The second theme arising from the analysis is the issue of professional judgement, which brings into question the education of teachers. Professional judgement is based upon a person's understanding of the nature and purpose of a mathematics assessment task grounded in an understanding of the nature and purpose of mathematics and mathematics education. Many studies (e.g., Ellis, Miller-Reilly & Pfannkuch, 1998) on teacher development aimed at improving or changing teaching state that knowing about new teaching strategies is not sufficient. An important ingredient, among others, is that teachers must discuss or study the theoretical basis or rationale for the new teaching method. The same criterion should apply to the introduction of new assessment strategies.

Bodin (1993) refers to the notion of professional test developers supplying teachers with good assessment devices that will "lead them to a better understanding of the knowledge of each student" (Bodin, 1993, p. 140). Whilst agreeing that assessment tasks would be improved greatly by publishers and test developers understanding and taking cognisance of the research in mathematics education, this research suggests that unless teachers understand the theoretical underpinnings of the assessment they will continue to choose, modify and alter such tasks to cater for their perception of mathematics and for their perception of their students. At the initial stages of implementing the curriculum there appears to have been a failure in addressing the theoretical knowledge associated with the introduction of new mathematics assessment strategies and with mathematics education itself. This is a system failure, not a teacher failure.

Warren and Nisbet (1999) indicate that a relationship between primary teachers' beliefs about the uses of assessment and their actual assessment practice exists. My interviews with these teachers tends to confirm such a relationship. However the similarity in beliefs about assessment and in the practice of assessment and also what is shared and not shared amongst teachers suggest that the school system enculturation process formulates how teachers operate in the school environment. In this 1995 "snapshot" of assessment in mathematics education, one learns that teachers have responded to the requirements of the curriculum by assessing students using a variety of assessment strategies and through learning objectives. The adoption of a mastery-based assessment approach has been the requirement of the system to enforce accountability. This means that the original aims of the writers of the mathematics curriculum to provide an integrated mathematics experience for students, which would allow higher order thinking skills to flourish, probably has been thwarted. But also perhaps thwarted is the politicians' aim to improve accountability, since the teacher in the system makes her own judgements. The question remains as to what the 2005 snapshot will reveal. Will teachers move away from objective-based assessment practice? Will students be experiencing an integrated mathematics assessment approach? Or will national or regional tests be driving assessment in the classroom?

References

- Bjorkqvist, O. (1997). Some psychological issues in the assessment of mathematical performance. In E. Pehkonen (Ed.), *Proceedings of the 21st annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 3-17). Lahti, Finland: Program Committee.
- Bodin, A. (1993). What does to assess mean? The case of assessing mathematical knowledge. In M. Niss (Ed.), *Investigations into assessment in mathematics education* (pp. 113-141). Dordrecht: Kluwer Academic Publishers.
- de Lange, J. (1995). Assessment: No change without problems. In T. Romberg (Ed.), *Reform in school mathematics and authentic assessment* (pp. 87-172). New York: State University of New York Press.
- Ellis, J., Miller-Reilly, B., & Pfannkuch, M. (1998). Ways in which some women came to influence mathematics education in New Zealand through the Equals network 1985-1995. In C. Keitel (Ed.) *Social justice and mathematics education* (pp. 220-237). Berlin: Freie Universitat Berlin.

- Kyriakides, L. (1998). Professional influences on teachers' perceptions of teaching and assessment in mathematics. In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 151-158). Stellenbosch, South Africa: Program Committee.
- Ministry of Education (1992). *Mathematics in the New Zealand curriculum*. Wellington, NZ: Learning Media.
- Niss, M. (1993). Assessment in mathematics education and its effects: An introduction. In M. Niss (Ed.), *Investigations into assessment in mathematics education* (pp. 1-30). Dordrecht: Kluwer Academic Publishers.
- Philippou, G., & Christou, C. (1996). Teachers' conceptions about mathematical assessment. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 131-138). Valencia, Spain: Program Committee.
- Romberg, T., Zarinnia, E. A., & Collis, K. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 21-37). Washington DC: American Association for the Advancement of Science.
- Thomas, G. (1998, June). *Pre-service training and professional development of teachers in mathematics: Where to from here?* Paper presented at the mathematics education research seminar, Ministry of Education, Wellington, New Zealand.
- Warren, E., & Nisbet, S. (1999). The relationship between the purported use of assessment techniques and beliefs about the uses of assessment. In J. Truran & K. Truran (Eds.), *Making the difference* (Proceedings of the 22nd annual conference of the Mathematics Education Research Group of Australasia, pp. 515-521). Sydney: MERGA.
- Watson, A. (1998). What makes a mathematical performance noteworthy in informal teacher assessment? In A. Olivier & K. Newstead (Eds.), *Proceedings of the 22nd annual conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 169-176). Stellenbosch, South Africa: Program Committee.¹

¹ *Acknowledgement.* I would like to thank Tamsin Meaney for her analyses of the interviews and helpful discussions concerning this paper.