Primary Teachers' Beliefs and Practices with Respect to Compulsory Numeracy Testing

Steven Nisbet
Griffith University
<s.nisbet@griffith.edu.au>

Peter Grimbeek
Griffith University
<p.grimbeek@griffith.edu.au>

This paper reports the results of a factor analysis of data from a survey of teachers' beliefs and practices relating to the compulsory numeracy tests conducted in Years 3, 5 and 7. The resulting six factors related to giving feedback to students, using the tests for diagnosis of pupils and content, changes in teachers' practice, comparing results with other schools, test validity, and preparing pupils for the tests. Analysis of factor scores showed the significance of professional development in teachers' practices associated with the tests. School location and size also had an effect on teachers' beliefs and practices associated with the tests.

Over the last 10 years, numeracy skills have been subject to much debate and scrutiny, resulting in increased pressure being placed on primary schools to improve outcomes and report on progress. A review of the school curriculum (Wiltshire, McMeniman, & Tolhurst, 1994) lead to the introduction of the *Year 2 Diagnostic Net* and *Year 6 Test* in schools in the mid 1990s (Queensland Schools Curriculum Council, 1996). Although the Year 6 Test was discontinued in 1997 (making way for the federally-initiated Year 3, 5 and 7 Tests), the Year Two Net continues to be used. It has been received well by primary teachers and has had a positive impact on their teaching of mathematics (Nisbet & Warren, 1999).

Further, at a national level, performance-based assessment and reporting was promulgated in the mid 1990s (Australian Education Council, 1994a), and all states were given individual responsibility for implementation of these procedures. Consequently in Queensland, Student Performance Standards (Australian Education Council, 1994b), were unsuccessfully introduced system-wide with teacher opposition, despite substantial funds provided for professional development of teachers (Nisbet, Dole & Warren, 1997).

In 1997, a National Literacy and Numeracy Plan was adopted in all states to (i) identify students at risk, (ii) conduct intervention programs, (iii) assess all students against national benchmarks, and (iv) introduce a national numeracy reporting system (Department of Education, Training and Youth Affairs, 2000). Consequently, annual compulsory state-wide testing was introduced for students in Years 3, 5 and 7 in 1998. In August each year, all students in Years 3, 5 and 7 in Queensland government schools sit for tests in numeracy.

In the Queensland tests, a broad interpretation of *numeracy* is assumed, embracing the perspectives offered by Willis (1998) that numeracy (i) includes concepts, skills and processes in mathematics, (ii) is described in terms of everyday situations in which mathematics is embedded, and (iii) implies that students can choose and use mathematical skills as part of their strategic repertoire. Hence the Queensland tests cover number, measurement, geometry, chance and data, and test skills of calculation (written, mental & calculator methods), and real-world problem solving.

A review of the Year 3, 5 and 7 testing program (Queensland School Curriculum Council, 1999) identified potential benefits and concerns related to such state-wide testing. The suggested benefits for teachers include the identification of students' strengths and weaknesses, data to inform planning and teaching, the provision of results for various groups (boys, girls, students of non-English speaking backgrounds, & Indigenous students), and identifying teachers' professional development needs. Issues of concern include narrowing the curriculum, a tendency to teach to the test, having assessment items not based on the classroom program, and the potential for misuse of results (e.g. the publication of 'league tables' of 'good' and 'bad' schools).

The reports sent to schools after the annual tests contain extensive information on the results of the tests for the school including: results for each test item and each section (number, space, measurement & data) for each year-level, for each subgroup (boys, girls, NESB, & indigenous students), and for each student, with comparisons with the state averages. Further, all incorrect answers are recorded for each item for each student, and items for which the school scored 15% above and 15% below the state average are listed. With such information supplied, teachers and administrators are in a position to identify strengths and weaknesses of the school's program, compare their results with those of other schools, and take what they may consider to be appropriate action.

The nature and extent of the action taken by schools varies across the state, and some of this information has been gathered by QSCC (later QSA) in surveys of participating schools. For example the survey undertaken in relation to the reports about the 2001 tests indicated that schools would make extensive use of the information in the reports. For instance, 80% of schools indicated that they would use the data for diagnosis of individual students' needs, and 78% indicated they would use the data to inform school programming.

However, it is not known whether these intentions reflect the opinion of class teachers (and not just the principal) and whether the schools and their teachers actually put the test results to such uses. Evidence gathered in a pilot study suggests that although schools may have good intentions, they don't actually get around to using the results. The current study was designed to determine the extent to which schools analyse and use the test data and teachers' views of the Year 3, 5 and 7 tests.

The adoption of the Year 3, 5 and 7 Numeracy Tests has been yet another change that primary teachers in Queensland have had to cope with in recent times. Much of the literature on teacher change and professional development acknowledges the importance of teacher beliefs as well as teacher knowledge in the cycle of professional growth. For instance, the importance of teachers' knowledge and beliefs in the cycle of professional growth was confirmed by Kyriakides (1996) who found that the failure of a mathematics curriculum change in a centralized system was due to the fact that teachers' perceptions of mathematics were inadequately considered at the adoption and implementation stages. Similarly, Philippou and Christou (1996) noted that if new ideas are to find their way into mathematics classrooms, it is imperative that change agents have a deeper understanding of classroom teachers' views, beliefs, conceptions and practices. Their study found that although teachers may be aware of and accept contemporary ideas (in their case about assessment), there can be a distance between their knowledge and intentions on the one hand, and their actual practice on the other hand.

The traditional model of implementing innovation assumes that teacher change is a simple linear process: staff development activities lead to changes in teachers' knowledge, beliefs and attitudes, which, in turn, lead to changes in classroom teaching practices, the outcome of which is improved student learning outcomes (Clarke & Peter, 1993). Later models of teacher change recognise that teacher change is a long term process (Fullan, 1982) and that the most significant changes in teacher attitudes and beliefs occur *after* teachers begin implementing a new practice successfully and can see changes in learning (Guskey, 1985). The professional development (PD) models of Clarke (1988) and Clarke and Peter (1993) are refinements of the Guskey model which recognise the on-going and cyclical nature of PD (focussing on knowledge, attitudes & beliefs) and teacher change.

Such models can help explain why some educational innovations are successful, and others not. The introduction of the Year 2 Diagnostic Net was successful because teachers saw positive outcomes for pupils and they valued the Net's overall effect (Nisbet & Warren, 1999). However the introduction of Student Performance Standards in mathematics was a failure because teachers did not believe that the extra work entailed in performance-based assessment and reporting was worthwhile. Further, they received little support for the move (Nisbet, Dole & Warren, 1997).

After five years of administration of the Years 3, 5 and 7 Tests it was considered appropriate to investigate the impact of the tests on schools. Hence the current study was devised. The aim of the study was to investigate teachers' attitudes to and beliefs about the Year 3, 5 and 7 tests (agreement with tests, & their validity & purposes), how schools and teachers use the test results (identifying students with difficulties & gaps in the curriculum), the impact of the tests on teachers' practices (preparation for the test, influence on content & method), and the responses of teachers and pupils to the tests. A further aim was to determine the effect of school location, school size, experience and extent of PD on such attitudes, beliefs and practices.

Methodology

A survey of teachers across Queensland government primary schools was conducted in 2003. A questionnaire was constructed containing items about teachers' attitudes, beliefs and practices relating to the Year 3, 5 and 7 Tests, plus background variable items relating to the teachers' grade level, teaching experience, school location, school size, and amount of professional development in mathematics teaching.

A sample of 56 primary schools representative of size, disadvantaged-schools index and geographical location across Queensland was selected[1] and a total of 500 questionnaires were sent to the schools (having estimated the number of teachers in each school from the data on pupil numbers). Although the response rate was small (121 responses i.e. 24.2%), the sample was representative of teachers' year level and position (Year 1 to Year 7, principal, deputy, & mathematics coordinator), teaching experience (from 1 year to 40 years), geographical location (capital city, provincial city, rural & remote), and school size (categories from <20 pupils to >400 pupils).

A first analysis of the data was an examination of the levels and spread of opinion with respect to the items, and the effects of background variables on response to the items. The effects of geographic location, size, and level of teaching were investigated by conducting chi-square tests on cross-tabulations of the substantive items with categories of location, size and teaching level. The results of the first analysis are reported elsewhere (Nisbet, submitted, 2004) and are summarised briefly in the 'results' section below.

Further statistical analysis was deemed appropriate in order to determine clearer patterns in the data in terms of clustering of questionnaire items and the effects of background variables such as geographical location, school size and teacher experience. Hence, the 29 items were entered into a factor analysis with the aim of identifying a set of conceptually relevant latent variables. Later, factor scores were saved to examine the relationships between specific background variables (geographical location, school size, teacher experience, & extent of PD).

Results

Looking at the survey results globally (Figure 1), it is clear that there are a number of issues about which teachers have strong views. For example, more than 80% of participants agreed or strongly agreed with statements relating to preparing students for the tests (how to fill in answers), and conducting practice tests. They also agreed or strongly agreed with the assertion that the test results arrived too late in the year to be of use. Fewer than 20% of participants agreed or strongly agreed to statements related to the tests being an indication of the teacher's ability or the school program, statements about pupils coping with the tests, the use of the tests for accountability purposes, checking class progress, planning teaching, or obtaining advice to analyse the schools results.

Figure 1. The percentage of participants agreeing or strongly agreeing to each of 29 items.

Looking at the responses item by item, the following observations can be made:

Most teachers don't believe the tests are valid indicators of pupils' numeracy ability, the teacher's ability, or the school's numeracy program.

Most teachers don't believe the tests ensure accountability or assist pupils' learning.

Only 47% of teachers agree with the tests in principle.

Less than half of the teachers make use of the results of the tests to identify pupils with difficulties, or to inform their planning.

The majority of teachers believe that the test results arrive too late to be of use.

The majority of teachers prepare their pupils for the tests, but few say that the tests influence what they teach or how.

The majority of teachers report that their pupils become anxious with the tests., whilst a minority report that their pupils cope with the tests.

School location had a significant effect on eight items – use of tests to identify gaps in content, pupils anxiety, agreement with tests, tests to ensure accountability, using results to identify difficult topics, reporting results to the community, and the influence of the tests on teaching (what & how) and assessing mathematics.

School size had a significant effect on four items – tests as a means of ensuring accountability, and the influence of the tests on teaching (what & how) and assessing mathematics.

Teaching experience correlated significantly with two items, namely, use of the test results to identify pupils with difficulties and use of the test results to plan teaching.

The next level of analysis concerns the clustering of items and the effects of background variables on revealed factors. To implement this, the 29 items relating to beliefs, attitudes and practices were entered into a factor analysis with the aim of identifying a set of conceptually relevant latent variables. The 29 items were factorable (KMO>0.800). However, the PAF analysis and Varimax (orthogonal) rotation produced a six-factor solution that was neither simple (some items loaded >.30 on more than one factor) nor interpretable (items did not group sensibly). After removal of items with loadings exceeding 0.30 on two or more factors, in an iterative process, a refined analysis with 15 of the 29 items resulted in a factorable (KMO=.766) six-factor solution that was both simple and highly interpretable. (See Table 1.)

Table 1.
Factor Solution for 15 Items (PAF extraction, Varimax rotation, ≧0.2 loadings shown)

Items/factors	1	2	3	4	5	6
Feedback on strengths	**.913**					
Feedback on weaknesses	**.813**					
Encourage students	**.706**	.203				
ID gaps in content		**.846**				
ID topics		**.837**				
ID pupil difficulties	.239	**.761**				
Influence how	.200		**.882**			
Influence assessment			**.807**			
Influence what			**.724**			
Comparing schools				**.883**	.236	
Compare with state				**.821**		
Teacher's ability			.224		**.798**	
School program				.279	**.728**	
Preparation						**.790**
Practice tests						**.783**

As shown in Table 1 above, the factor analysis resulted in a highly interpretable six-factor solution (with N of factors set arbitrarily), such that the six factors could be labeled and described as follows (listed in order of factors):

- Feedback (three items): Teachers using the test results to encourage students, and to give them feedback on their strengths, and weaknesses.
- Diagnosis (three items): School using results for diagnostic purposes, i.e. to identify pupils with difficulties, identify gaps in content, and identify topics causing difficulties.
- Teacher change (three items): Tests influencing teachers' practice in mathematics – what and how they teach it, and how they assess it.
- Comparison (two items): Tests being a good way of comparing the school with the other schools and the whole state.
- Validity (two items): The tests being seen as valid indicators of the teachers' ability and the school's numeracy program.
- Preparation for tests (two items): Teachers showing pupils how to fill in answers, and giving practice tests.

Factor scores for each of the six factors were saved (Anderson-Rubin method: Producing orthogonal z score values) to examine the locus of relationships between specific background variables and these six factors. The four background variables of interest were as follows: (i) Geographical Location, (ii) School Size, (iii) Teacher Experience, and (iv) Amount of PD. In order to facilitate the multivariate analysis, these background variables were collapsed in such a way that an empirical examination of interactions became viable, that is, so that there were three or more scores per cell. Also, for this reason, the effect of geographical location was examined separately.

The first MANOVA considered the effect of School Size (1-400; >400 students), Teacher Experience (0-14 years; more than 14 years), and Amount of PD (none; some) on the six labeled factor scores (Feedback, Diagnosis, Teacher Change, Comparison, Validity, and Preparation). The analysis revealed significant multivariate effects (Pillai's Trace, $p < 0.05$) for School Size, PD, Teacher Experience by School Size, and PD by School Size.

Examination of univariate effects for School Size significantly predicted the likelihood of Teacher change ($F(1, 117) = 5.759$, $p < 0.05$). Follow-up examination of marginal means indicated that participants from smaller schools (1-400 students) responded significantly more positively than others, i.e. the tests influenced them more in what mathematics they taught, how they taught and assessed it.

Examination of univariate effects for Amount of PD significantly predicted the likelihood of Diagnosis ($F(1, 117) = 6.208$, $p < 0.05$) and Preparation ($F(1, 117) = 4.456$, $p < 0.05$). Follow-up examination of marginal means indicated that participants with some exposure to PD responded significantly more positively than others in terms of both using the tests for diagnosis of pupils, gaps in content and difficult topics, and preparing the students for the tests.

Examination of univariate effects for the two-way interaction between School Size and Teacher Experience significantly predicted the likelihood of Comparison ($F(1, 117) = 6.436$, $p < 0.05$) and Preparation ($F(1, 117) = 6.316$, $p < 0.05$). Follow-up examination of the marginal means for both Comparison and Preparation indicated that whereas participants with 0-14 years of teaching experience responded significantly more positively if they were from smaller schools (1-400 students), those with more years of teaching experience responded significantly more positively if from larger schools (401+ students).

Examination of univariate effects for the two-way interaction between School Size and Amount of PD significantly predicted the likelihood of Feedback ($F(1, 117) = 4.055$, $p < 0.05$), Teacher Change ($F(1, 117) = 4.231$, $p < 0.05$), and Comparison ($F(1, 117) = 5.405$, $p < 0.05$). Follow-up examination of the marginal means for Feedback and Comparison indicated that participants with who had been exposed to some PD in mathematics were more positive in their responses if from smaller schools and the reverse for larger schools. In terms of Teacher Change, those with PD were more likely to be positive if from larger schools and the reverse for smaller schools.

What the above highlights is the importance of School Size in relation to the positivity of participant responses. More generally, the significant two-way interactions suggest that Teacher Experience and exposure to PD has a positive effect in smaller schools but not in larger schools. The only dependent variable positively influenced by School Size in this respect was Teacher Change. That is, teachers in larger schools with some exposure to PD were more likely to have reported a greater influence of the tests on their practice.

The second MANOVA examined the effect of Geographical Location on the six dependent variables outlined above. The analysis revealed a significant multivariate effect for Geographical Location (Pillai's Trace, $p<.05$).

Examination of univariate effects for Geographical Location significantly predicted the likelihood of Teacher Change ($F(1, 117) = 3.264$, $p < 0.05$) and Validity effects ($F(1, 117) = 3.427$, $p < 0.05$). Follow-up examination of marginal means indicated that participants from rural and remote locations responded significantly more positively than those from other locations. In other words, teachers in rural and remote schools were more likely to have reported a greater influence of the tests on their practice, and their perception of the tests as valid indicators of their ability and the school's numeracy program. Table 2 provides a summary of the effects of background variables on factor scores.

Table 2.
Summary of Effects of Background Variables on Factor Scores

Factors/Variables	Geographical Location	Teacher Experience	School Size	Amount of PD
1. Feedback			Negative interaction	
2. Diagnosis				Positive effect
3. Teacher Change	Positive effect for rural and remote		Negative effect	
			Positive interaction	
4. Comparison		Positive interaction		
			Negative interaction	
5. Validity	Positive effect for rural and remote			
6. Preparation				Positive effect
		Positive interaction		

Conclusions

The initial analysis revealed that although there is a great diversity of beliefs and practices among teachers, attitudes to the tests appear to be very negative. It is of concern that the tests have not greatly influenced teaching practices and that the results of the tests are not being used to any great extent to inform planning apart from identifying gaps in the schools' mathematics programs.

A number of conclusions can be made from the results of the factor analysis. Firstly, the teachers who scored the highest level on test validity (those in rural & remote schools), were also the teachers who were more likely to change their practice in the light of the results of the tests. This supports Clarke and Peter's (1993) recognition of the important role of teachers' attitudes and beliefs in the process of changing teachers' practices.

Secondly, large schools (>400 students) seem to have a distinct perspective on the tests compared to other schools: their more experienced teachers are more likely to change their practice in the light of test results, and see the tests as a valid way of comparing their schools with similar schools. It seems that these schools and teachers are very competitive and consequently place more emphasis on preparing the students for the tests in order to maximise the scores of their students.

Thirdly, professional development is a key element in any discussion of teacher change and teachers' practice. Exposure to professional development was a significant variable in five of the six factors identified – feedback, diagnosis, teacher change, comparison and preparation – either as a single variable or in interaction with school size. This result confirms the strong claim by Aichele and Coxford (1994) that the professional development of mathematics teachers has the power to transform the entire field of education. In this study, professional development exhibited more influence on teachers' practices than teacher experience, which was a significant variable in only two factors – comparison and preparation – and then it was in interaction with school size. Professional development must be recognised for its potential for changing teacher practice significantly and for the better, and its important role in improving numeracy outcomes in schools.

[1] Staff from Australian Council for Educational Research (ACER) provided assistance with the sample design and selected the sample of schools.

References

Aichele, D., & Coxford, A. (1994). *Professional development for teachers of mathematics (1994 Yearbook).* Reston, Va: National Council of Teachers of Mathematics.

Australian Education Council (1994a). *Mathematics work samples.* Carlton, Vic: Curriculum Corporation.

Australian Education Council (1994b). *Student performance standards in mathematics for Queensland schools.* Carlton, Vic: Curriculum Corporation.

Clarke, D. J., & Peter, A. (1993). Modelling teacher change. In B. Atweh, C. Kanes, M. Carss, & G. Booker (Eds.), *Contexts in Mathematics Education* (Proceedings of the 16th annual conference of the Mathematics Education Research Group of Australasia, Brisbane, pp. 167-175). Brisbane: MERGA.

Clarke, D. J. (1988). Realistic assessment. In D. Firth (Ed.), *Maths counts – who cares?,* Parkville, Victoria: Mathematical Association of Victoria, pp. 187-192.

Department of Education, Training & Youth Affairs (2000). *Numeracy, a priority for all.* Canberra, ACT: Commonwealth of Australia.

Fullan, M. (1982). *The meaning of educational change.* New York: Teachers College Press.

Guskey, T. (1985). Staff development and teacher change. *Educational Leadership, 42*(7), 57-60.

Kyriakides, L. (1996). Primary teachers' perceptions of policy for curriculum reform in Cyprus with special reference to mathematics. *Mediterranean Journal of Educational Studies, 1*(2), 77-93.

Nisbet, S. (submitted). The impact of state-wide numeracy testing on the teaching of mathematics in primary schools. Paper submitted to 28th Annual Conference of the International Group for the Psychology of Mathematics Education. Bergen, July 2004.

Nisbet, S., & Warren, E. (1999). The effects of a diagnostic assessment system on the teaching of mathematics in the primary school. In O. Zaslavsky (Ed.), *Proceedings of 23rd Annual Conference of the International Group for the Psychology of Mathematics Education (*Vol. 3, pp. 337-344). Haifa: PME.

Nisbet, S., Dole, S., & Warren, E. (1997). Cooperative professional development for mathematics teachers: A case study. In F. Biddulph & K. Carr (Eds.). *People in Mathematics Education* (Proceedings of the 20th Annual Conference of MERGA, Rotorua, pp. 369-375) Sydney: MERGA.

Philippou, G., & Christou, C. (1996). Teachers' conceptions about mathematical assessment. In L. Puig & A. Gutierrrez (Eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education PME 20 (Vol. 4),* Valencia, Spain, July 8-12, 1996 pp. 131-138.

Queensland Schools Curriculum Council (1996). *The year 2 Diagnostic Net Handbook for Schools – Shaping the Future.* Brisbane The Queensland Government Press: Brisbane.

Queensland School Curriculum Council (1999). *Review of Queensland Literacy and Numeracy Testing Programs, 1995-1999: Issues Paper.* Brisbane: Queensland School Curriculum Council.

Willis, S. (1998). Which numeracy? *Unicorn, 24* (2), 32-42.

Wiltshire, K., McMeniman, M. & Tolhurst, T. (1994). *Shaping the future: Review of the Queensland school curriculum.* Brisbane: Queensland Legislative Assembly.