

# THE IMPACT OF STATE-WIDE NUMERACY TESTING ON THE TEACHING OF MATHEMATICS IN PRIMARY SCHOOLS

Steven Nisbet

Griffith University

*This paper reports on teachers' views of the effects of compulsory numeracy testing in Years 3, 5, and 7 in Queensland schools. Teachers were surveyed on (i) the validity and worth of the tests, (ii) the impact the tests had on their teaching of mathematics, and (iii) how they were using the results of the tests. Although the results reveal a great diversity of beliefs and practices among teachers, attitudes are very negative. The tests have not greatly influenced teaching practices and 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.*

## INTRODUCTION

In Queensland (Australia) primary schools have experienced an increased emphasis on numeracy (and literacy) skills since the mid 1990s. A review of the school curriculum (Wiltshire, McMeniman, & Tolhurst, 1994) led 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 proposed Year 3, 5 & 7 Tests), the Year 2 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 its implementation. Consequently in Queensland, Student Performance Standards were introduced but faced teacher opposition and so was unsuccessful, despite the fact that substantial funds were provided for professional development projects (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 & 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 schools sit for tests in numeracy (and literacy). The tests are devised by the Queensland Studies Authority (QSA) and are distributed to all government schools.

In Queensland, 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 & 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 naturally 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 & 7 tests.

The adoption of the Year 3, 5 & 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 curriculum 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 student learning (Guskey, 1985). The professional development models of Clarke (1988) and Clarke and Peter (1993) are refinements of the Guskey model which recognise the on-going and cyclical nature of professional development (focussing 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).

The primary aim of the current study was to investigate (a) teachers' attitudes to and beliefs about the Year 3, 5 & 7 tests (agreement with tests, & their validity & purposes), (b) how schools and teachers use the test results (identifying students with difficulties & gaps in the curriculum), (c) the impact of the tests on teachers' practices (preparation for the test, influence on content & method), and (d) the responses of teachers and pupils to the tests. A secondary aim was to determine the effect of school location, school size, and level of teaching on the above attitudes, beliefs and practices.

## **METHODOLOGY**

This study was conducted by survey method. A questionnaire was constructed containing items about teachers' attitudes, beliefs and practices relating to the Year 3, 5 & 7 Tests (as described above), plus items relating to the teachers' grade level, teaching experience, school location and school size, and an item for 'any other comments'. The results of a pilot study of 34 teachers in city and rural schools

conducted in the previous months (Nisbet, 2003) were used to revise and expand the questionnaire items. A five-point Likert scale (from 1 = 'disagree strongly' to 5 = 'agree strongly') was provided for responses, and teachers were also invited to comment in selected items. A sample of 56 primary schools representative of size, disadvantaged-schools index and geographical location across Queensland was selected<sup>1</sup> 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 (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 (Brisbane i.e. capital city, provincial city, rural & remote), and school size (categories from <20 pupils to >400 pupils).

## RESULTS

The results are presented in six sections – (i) teachers' attitudes to and beliefs about the Year 3, 5 & 7 tests (ii) how schools and teachers use the results of the tests, (iii) the impact of the tests on teachers' practices, (iv) the responses of teachers and pupils to the tests, (v) the effect of school location and size, teaching level and teaching experience on the above attitudes, beliefs and practices, and (vi) other comments.

### **Teachers' attitudes to and beliefs about the Year 3, 5 & 7 tests**

Opinion is divided on agreement with tests in principle, however the overall view of the tests and their purposes is quite negative. Although more teachers agree with the tests (47.1%) than disagree (33.8%), almost one in five (18.2%) are undecided. A minority of teachers agree that the tests are a good way of ensuring accountability (17.3%), a good way of comparing their school with other schools (34.7%), or a good way of comparing their school with the state (38%). Further, the majority of teachers (60.3%) think that the tests do nothing to assist pupils' learning.

Teachers think that the tests have little validity. Only 25.7% believe that the results of the tests give an accurate indication of the pupils' numeracy ability. Fewer teachers (12.4%) believe that the results of the tests give an accurate indication of the quality of the school's numeracy program. Fewer still (5.8%) believe that the results give an accurate indication of the teacher's ability to teach mathematics.

### **The uses that schools and teachers make of the results of the tests**

It appears that schools make use of the results in some ways but not others. For example, 67.5% of teachers report that their school analyses the results to identify topics causing difficulties. Further, 59.2% believe that their school analyses the results to identify gaps in content taught, and 66.2% believe that the school analyses the results to identify pupils experiencing difficulties. However, only 36.7% of teachers report that the school uses the results of the numeracy tests to notify parents about the school's overall performance. Similarly, only 38.4% of teachers report that the school informs the community about the school's overall performance, and 7.5% of teachers report that the school obtained expert advice on analysing the results.

At the personal teacher level, the data appear more negative compared to those above in regard to the school's use of the test results. For instance, only 40.5% of teachers report using the results to identify individual students who are having difficulties. Fewer teachers give students feedback on their strengths (38.9%) and weaknesses (27.2%), or use the results to encourage pupils (36.4%). Only 19.9% of teachers report using the results to judge how well the class is progressing, and only 21.5% use the results to plan their teaching. These low figures may be due to the fact that 77.7% of teachers believe that the results arrive too late in the year to be of use.

### **The impact of the tests on teachers' practices**

The level of impact of the tests on teachers' practices varies greatly depending on the domain of practice. The majority of teachers (91.7%) report showing pupils how to fill in the answers before the day of the test (e.g. colouring response bubbles, writing numbers in boxes) and giving pupils a practice test before the day (89.2%). However, very few report that the tests have influenced what they teach in mathematics lessons (31.7%), how they teach mathematics (26%), or how they assess it (20%).

### **The responses of pupils to the tests**

It appears, according to teachers, that the numeracy tests are a negative experience for pupils. Only 20.6% of teachers report that their pupils cope with the tests, and the majority of teachers (61.2%) report that their pupils become anxious with the tests.

### **The effects of geographical location, school size, level of teaching and experience**

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.

School location had a significant effect on eight (out of 29) items: Brisbane schools use the test results to identify gaps in the curriculum more than other schools ( $\chi^2 = 23.7, p = .02$ ). Pupils in Brisbane schools are more likely than other pupils to be anxious with the tests ( $\chi^2 = 41.8, p = .0001$ ). Teachers in Brisbane schools agree less than other teachers with the whole idea of the tests ( $\chi^2 = 31.5, p = .008$ ) and the idea that the tests ensure accountability ( $\chi^2 = 23.8, p = .02$ ). Rural schools are more likely to use the results to identify topics causing difficulties ( $\chi^2 = 33.6, p = .001$ ) and are more likely to report the test results to the community ( $\chi^2 = 21.6, p = .04$ ). Rural teachers more than other teachers are influenced by the tests in regard to how they teach mathematics ( $\chi^2 = 27.6, p = .006$ ) and how they assess it ( $\chi^2 = 32.2, p = .001$ ).

School size had a significant effect on only four items: Teachers in large schools (>400 pupils) agree less with the idea that the tests ensure accountability than teachers in other areas ( $\chi^2 = 47.01, p = .001$ ). Teachers in large schools are influenced less by the tests in regard to what mathematics they teach ( $\chi^2 = 44.4, p = .001$ ), how they teach it ( $\chi^2 = 47.8, p = .001$ ), and assess it ( $\chi^2 = 51.5, p = .001$ ).

Level of teaching had a significant effect on only five items. For instance, Year 3 teachers used the test results to identify pupils experiencing difficulties less than

teachers of other Year levels ( $\chi^2 = 10.1, p = .04$ ). Teachers who thought their pupils cope with the tests more than pupils in other grades were Year 6 teachers ( $\chi^2 = 9.7, p = .045$ ), and Year 7 teachers ( $\chi^2 = 13.6, p = .01$ ). Year 6 teachers agreed with the idea of testing more than other teachers ( $\chi^2 = 11.4, p = .045$ ). Lastly, principals believed more than the other teachers that the tests influenced what mathematics was taught in the school ( $\chi^2 = 9.8, p = .04$ ).

Years of teaching experience correlated significantly with only two items – positively with use of the test results to identify pupils experiencing difficulties ( $r = .24, p = .01$ ), and negatively with use of the test results to plan teaching ( $r = -.19, p = .04$ ).

### **Other comments**

Issues raised by teachers were mostly negative and related to the tests themselves (problems with the language of the tests, the ability of students especially those in Year 3 to read the tests, ambiguity and formatting of the tests), the way the tests are administered (time of year & lack of consistency across classes & schools), curriculum issues (a perceived mismatch between the tests & the syllabus, unsuitability to multi-age classrooms), the effect on pupils (emotional stress & reinforcement of negative self image, especially in Year 3), effect on teachers (stress), and philosophical concerns (unfairness of some schools using the actual test as a practice test beforehand).

## **DISCUSSION**

It is clear from the survey data that teachers have not embraced the Year 3, 5 & 7 Numeracy Tests to any great extent, nor have the tests had much impact on their teaching. Less than half of the teachers agree with the tests in principle, and the majority think that the tests do little to assist learning or ensure accountability. The data also reveal some inconsistencies in the teachers' responses, especially on the use made of the test results. Although the majority report that the school uses the results to identify topics causing difficulties and gaps in the curriculum, only 21.5% report that they personally use the results to inform their planning, and only 19.9% use the results to judge how well the class is progressing. Similarly, two-thirds of the teachers believe that the school analyses the results to identify pupils experiencing difficulties, but less than 40% give pupils feedback on their strengths or weaknesses.

There are three possible reasons for these inconsistencies. Firstly, the great majority believe that the results arrive at the school too late in the year to be of any use to the school or the individual teachers. The results arrive within two or three weeks of the end of the school year, a time when the Year 7 students are getting ready to transfer to high school, and the pupils in Years 3 and 5 are soon to move up to Years 4 and 6 respectively. At the start of the following year, their new teachers have had a holiday break and are then preparing for their new classes. Last year's test results are not high on the agenda. The issue of timing of the tests and the release of results should be seriously considered by the testing authority (QSA) with a view to giving schools more opportunity to take advantage of the data available in the schools' test results.

Teachers' attitudes to the tests may change for the better if the results arrive earlier, giving schools ample opportunity to analyse and act upon the results.

Secondly, it is clear that not many schools seek expert advice on how to analyse the test results and what action to take in the light of the analysis. There is also an inconsistency between the teachers' responses and the schools' intentions which are signalled in the reports sent to the testing authority. Intentions don't seem to match the reality afterwards. Although the test-result data sent to schools are comprehensive, schools may need expert assistance in analysing and interpreting the data, and secondly in working out what changes need to be made to the school work program, classroom pedagogy, and teaching/learning support.

Thirdly, the low degree of usage of the test results may be symptomatic of the teachers' negative attitudes and beliefs as expressed by a lack of support for the tests and a belief that the tests have little validity. This situation confirms the importance of beliefs and attitudes in teacher change (Guskey, 1985; Clarke & Peter, 1993).

In terms of the Clark and Peter (1993) model of professional growth, practical issues such as lateness of the reports and lack of expertise or support for analysing and interpreting the test data are probably limiting the amount of activity in the *Domain of Practice* which in turn is limiting teachers seeing valued outcomes (*Domain of Inference*) and influencing teachers knowledge and beliefs (*Personal Domain*). Thus the cycle of professional growth is severely impeded. It appears that the potential for enhancing numeracy outcomes and improving practice has not yet been fully tapped.

It seems that teachers perceive the Year 3, 5 & 7 Numeracy Tests quite differently to the Year 2 Net, which is seen in a very positive light (Nisbet & Warren, 1999). There is obvious concern expressed by most teachers about their students coping with the numeracy tests and the levels of anxiety caused by the tests, especially in Year 3. Such concerns about testing children in Year 3 give reason to investigate (a) the validity and appropriateness of putting children of a young age through such an experience, and (b) the alternative of extending the Diagnostic Net into Year 3.

The fact that the effects of school location, school size, teaching level, and years of experience were evident in only a few items implies that the problems and issues revealed in the data are system wide, and not confined to small pockets of schooling. Teachers' negative beliefs and attitudes to the numeracy tests are proving to be a barrier to their acceptance and limiting efforts to taking advantage of the tests to improve students' numeracy outcomes and attitudes. To turn this around, teachers need to see something positive in relation to the improvement of pupils' performance, and this in turn needs an 'external source of information, stimulus or support' (Clarke & Peter, 1993). This could take the form of an initial trial of a numeracy-enhancement project involving specific professional development for teachers and school administrators, publication of the benefits obtained through the trial, and a subsequent extension of the project into other schools and districts.

<sup>1</sup> Staff from Australian Council for Educational Research (ACER) provided assistance with the sample design and selected the sample of schools. The ACER sampling frame is compiled annually from data provided by the Commonwealth and each State and Territory education system.

## References

- 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 Sixteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA), Brisbane, July 9-13, 1993. Brisbane: MERGA. pp. 167-175.
- 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.
- Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching*, 15(1), 13-33.
- 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. (2003). Teachers' views on the effect of compulsory state-wide numeracy testing on the teaching of mathematics in the primary school. In B. Bartlett, F. Bryer & D. Roebuck (Eds.) *Reimagining Practice: Researching Change*. Proceedings of the 1<sup>st</sup> Annual International Conference on Cognition, Language and Special Education, at the Crowne Plaza Hotel, Surfers Paradise, Australia, 5-7 December 2003.
- 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 the 23<sup>rd</sup> Annual Conference of the International Group for the Psychology of Mathematics Education*. Haifa, Israel, July 25-30, 1999. Vol. 3, pp. 337-344.
- 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 Mathematics Education Research Group of Australasia (Rotorua, July 1997)*. Hamilton, Waikato, NZ: MERGA.
- Philippou, G., & Christou, C. (1996). Teachers' conceptions about mathematical assessment. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20<sup>th</sup> 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*. The Queensland Government Press: Brisbane.
- Queensland School Curriculum Council (1999). *Review of Queensland Literacy and Numeracy Testing Programs, 1995-1999: Issues Paper*. Queensland School Curriculum Council: Brisbane.
- Willis, S. (1998). Which numeracy? *Unicorn*, 24 (2), 32-42.
- Wiltshire, K., McMeniman, M. & Tollhurst, T. (1994). *Shaping the Future: Review of the Queensland School Curriculum*. Brisbane: Queensland Legislative Assembly.