Beyond the Basics: Improving Indigenous Students' Numeracy

Rosemary Callingham
University of Tasmania
< rosemary.callingham@utas.edu.au>

Patrick Griffin
University of Melbourne
<p.griffin@edfac.unimelb.edu.au>

Improving numeracy for Indigenous students is a high priority in Australian education. One approach is that taken by the INISSS (Improving Numeracy for Indigenous Students in Secondary Schools) project in Tasmania. The INISSS program was evaluated in part through measuring students' numeracy outcomes using specially developed performance assessment tasks. The results from this process show that the program was successful in meeting its goal of improved numeracy outcomes.

The underachievement of Indigenous students is a continuing issue of significance in Australia (Groome & Hamilton 1995; MCEETYA 2000). On all indicators reported through national reporting processes, including numeracy outcomes, Indigenous students lag behind non-Indigenous students. Recently, strategic initiatives have attempted to address this issue (McRae et al. 2000). Although many of these have been deemed successful, the measures used to determine success have not included students' learning outcomes.

This paper reports the findings from the evaluation of the first phase of a program aimed at Improving Numeracy for Indigenous Students in Secondary Schools (INISSS) (Nicholson 1999). The goal of the program was to improve the numeracy outcomes of all students but particularly those of Indigenous students – in other words to close the 'gap' between the achievement levels of the two groups. The INISSS program was a Tasmanian initiative in response to Indigenous students' poor numeracy achievement in statewide testing in 1997. It provided professional development to 40 teachers and Aboriginal Education Workers in 19 project schools. The professional development aimed to develop teachers' skills in teaching for numeracy in ways that appeared to support Indigenous students' learning. These included small group work and practical investigations (Hughes & More 1997; Groome & Hamilton 1995). To support this, Task Centre materials were used that had been developed in collaboration with Indigenous communities in Queensland (Williams 1997).

It has been argued for a number of years that assessment tools should match teaching strategies (Shepard 2000). It seemed inappropriate, therefore, to use a multiple-choice or short answer test as the evaluation tool for assessing students' numeracy outcomes when the approaches to teaching that were advocated in the program emphasised collaboration and investigation. For this reason several performance assessment tasks that allowed students to work together were developed as evaluation tools (Callingham 1999; Callingham & Griffin 2000).

Although performance assessments have been hailed as one way of improving assessment processes, they create a number of technical difficulties (Linn 1994). These issues need to be addressed if the use of performance assessments is to provide useful evidence for evaluation or other purposes. A central concern is the construct validity of the task (Messick 1994). Does the task measure what it sets out to, rather than some other, equally valid, but different outcome from that targeted? There is evidence that some types

of task do not assess 'valued' mathematical outcomes so much as 'valued' workplace practices (Moschkovich 1998). This issue was addressed in the assessment tasks' design by basing them on the kinds of mathematical materials being used in classroom. Although students were able to undertake the tasks in a variety of ways the focus was mathematical, and this was recognised in the scoring scales (Callingham & Griffin 2000).

A second issue is that relating to the consistency of marking. Teacher-marked tasks have been criticised for lacking common standards of comparison (Linn 1994). The INISSS tasks were accompanied by an analytical scale that addressed the quality of the mathematical thinking demonstrated by students, rather than generalised criteria. In this way it was hoped to combine the performance based approach to assessment with tight and reliable rating scales (Callingham & Griffin 2000).

A final issue is that of generalisability across tasks (Linn 1994). The situated nature of many performance assessments means that often a student's performance cannot be generalised across the measured domain and many different tasks may need to be used. This was addressed by using measurement techniques based on Rasch modelling (Griffin 1997) and a study design that allowed the tasks to be placed on a single scale of generalised outcomes (Callingham & Griffin 2000).

By addressing these issues at the design stage of the study, performance assessments were produced that could provide quality evidence about students' numeracy outcomes in those schools that were part of the project.

Research Questions and Methodology

The overall aim of this evaluation was whether or not the INISSS program had met its intended outcomes. There were, however, subsidiary research questions addressed as part of the study.

1. Is it possible to use complex performance assessment tasks to evaluate a program intending to enhance students' numeracy?
2. Did a program that provided professional development to teachers influence their students' numeracy outcomes?

To answer these questions, a pre-test/post-test model of evaluation was used with Year 8 students as the target group for evaluation. This group was chosen because most project teachers were teaching classes in the year group. All Year 8 students in the project schools, whether or not they were being taught by teachers who were part of the project, undertook two assessment tasks early in the school year of 1999 (A1), and a further two tasks at the end of 1999 (A2). An overlapping test design allowed all students and tasks to be placed on a single scale (Callingham & Griffin 2000).

In addition teacher and student questionnaires were administered. The student survey covered background information about the students and their attitudes to mathematics and to school and was given to all Year 8 students who undertook A2. This was based in part on an instrument that had been used with Indigenous students elsewhere (Godfrey, Partington, Haslett, Harrison & Richer 1998). The teacher survey asked about teaching and assessment methodology, and recent professional development experiences. All mathematics teachers of the Year 8 students completed this, whether or not they had

undertaken INISSS professional development. This paper reports on students' learning outcomes and some aspects of the teacher questionnaire.

Between assessment A1 and assessment A2, a period of six months, approximately 40 teachers and Aboriginal Education Workers in the project schools undertook a program of professional development focussing on improved pedagogy. This was provided at two-day residential sessions as well as single days. At each professional development session teachers undertook to trial some aspect of new teaching methodology in their classrooms, and then report on the outcomes at the next session. This program stopped at the end of 1999 and no further professional development was provided for teachers in these schools.

In 2000, the same student cohort, now in Year 9, undertook a third assessment process using two assessment tasks (A3). One of these, Street Party (Callingham & Griffin 2000), provided five unchanged questions to create an anchor to earlier administrations, although the rest of the task had been revised. The other was totally new. This assessment provided longitudinal data about students' outcomes. The data collection is summarised in Table 1. In each case, teachers administered the assessment tasks in their classrooms and marked their own students' work (Callingham & Griffin 2000).

Table 1
INISSS Data Collection

Assess-ment	Indigenous students	Non-Indigenous students	Total students	Date	Grade	Additional information
A1	142	1700	1842	Apr 1999	Year 8	
A2	130	1567	1697	Oct 1999	Year 8	Teacher and student surveys
A3	96	1535	1631	Oct 2000	Year 9	

The data were analysed using the *Quest v.2.1* computer program (Adams & Khoo 1996), which provided Rasch analyses that allowed the relative positions of students and items to be plotted on a single scale. The five unchanged items were used to create an anchor file using item difficulties from the first assessment process. This allowed each assessment administration to be scaled through using a process of common item linking to the initial analysis. In this way a comparison of student achievement over the three assessment periods could be obtained (Griffin 1997).

Findings

Student Performance

Not surprisingly, the mean score for both Indigenous and non-Indigenous students rose over the three assessment periods. The mean logit scores are summarised in Table 2, together with the standard error of measurement. The larger standard errors associated with the scores for Indigenous students can be attributed to their relatively small numbers.

Table 2
Mean Logit Scores and Standard Errors for INISSS Assessments

	A1	A2	A3	SE A1	SE A2	SE A3
Non Indigenous students	-0.34	0.14	0.46	0.03	0.03	0.05
Indigenous students	-0.66	-0.24	0.26	0.09	0.10	0.10

Figure 1 shows the summary results for each of the three assessment events for Indigenous and non-Indigenous students in graphical form. It is noticeable that the 'gap' between the outcomes for the two groups narrowed between A2 and A3. This result suggested that the INISSS program had met its intended outcome of improving numeracy for all students, but particularly Indigenous students.

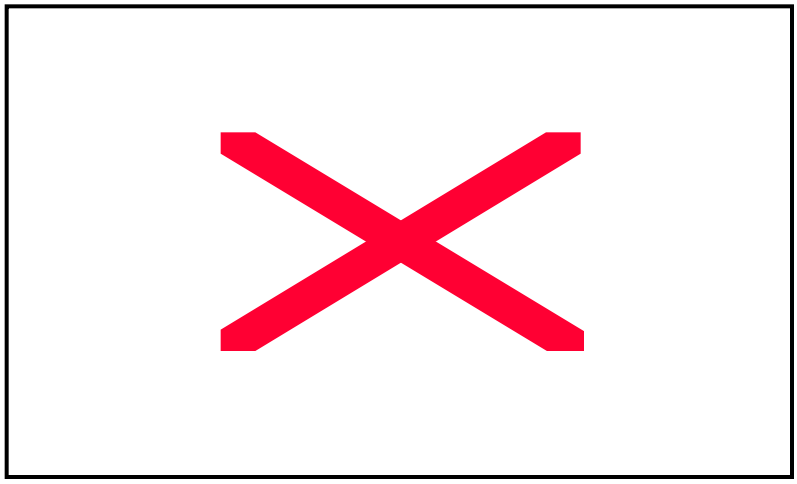

Figure 1. Mean logit scores from INISSS assessments.

When these results were further broken down by sex, some surprising detail emerged. This is shown in Figure 2. This indicates that the improvement of Indigenous students between A2 and A3 was due in large part to the performance of Indigenous girls, which approached that of non-Indigenous girls in A3. The generally better performance of girls in these assessments is in keeping with other research that suggests that girls perform better in assessment events that require writing (e.g. Lokan 1999).

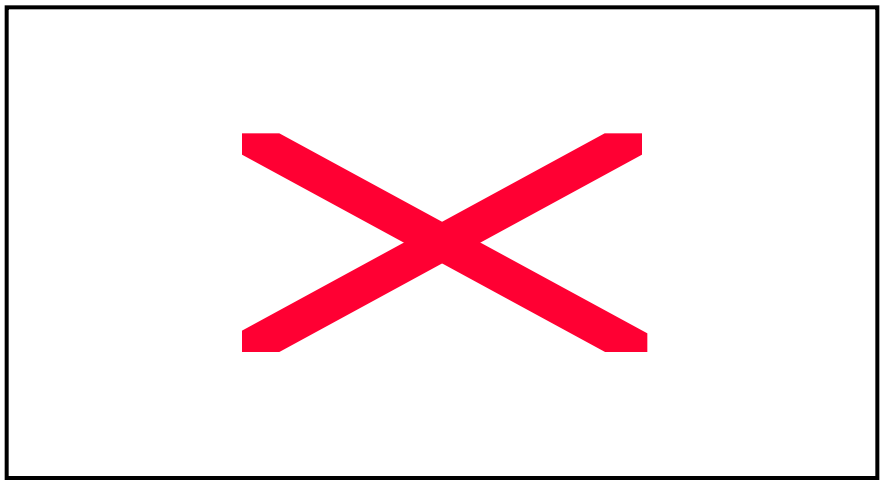

Figure 2. INISSS results by Indigenous status and sex.

School Results

When the findings were broken down by school, not every school showed improvement. Figure 3 shows the results by school for each of the three assessments.

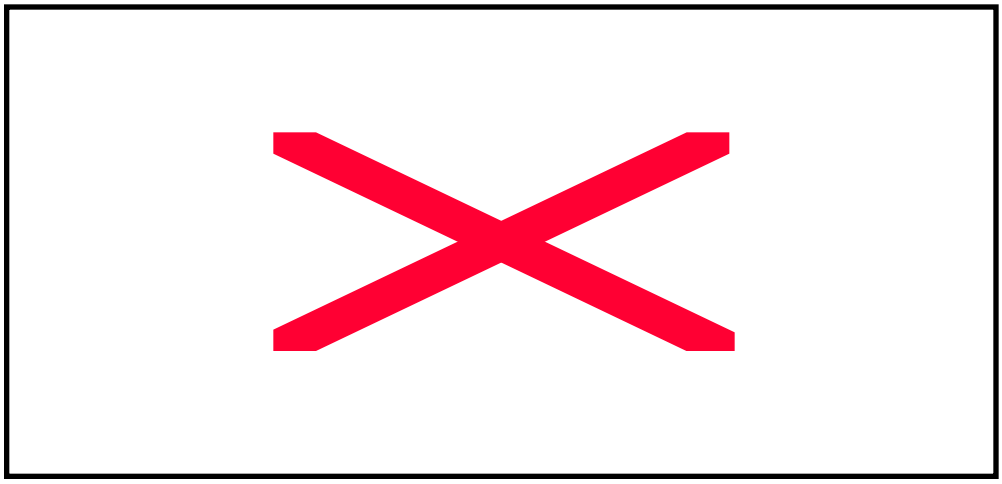

Figure 3. INISSS results by school.

Although most schools showed improvement, school *a*, for example, went backwards. School *k* had a high mean score initially, and made very little change, whereas school *l* made consistent large gains. Schools *h*, *p* and *r*, although they continued to be among the poorest performing schools in A3, also made big improvements. These results suggest that conditions within the schools contributed to the overall results, in keeping with other research about school effects (Hill, Rowe, Holmes-Smith & Russell 1996).

How could these differences be explained? Since teachers marked their own students' work, one possible cause of difference was inconsistent marking patterns across schools. This was addressed by considering the fit to the Rasch model. Fit measures are obtained by

comparing the observed responses with those predicted by the model. Systematic misfit within a school could suggest that the teachers in that school were applying different interpretations of the scoring rubrics. The fit across schools is shown in Figure 4 as the mean *infit* value. The horizontal lines at 0.7 and 1.3 logits show the acceptable limits of this measure of fit (Adams & Khoo 1996).

From this it can be seen that there is only one instance of serious misfit – school *e* in A3. This suggests that the issues of consistency of teacher judgement were addressed by providing a tight and rigorous analytical scoring rubric, and that teachers were able to use this consistently across all assessment events.

A contributing factor to the differences among schools could have been the degree of commitment to the program by the teachers and school management. Examination of records of attendance at the professional development sessions, showed that school *a* had intermittent attendance by a number of different teachers. Schools that had made large gains, however, such as schools *h*, *l* and *p*, were characterised by a strong team approach. At least two teachers regularly attended every professional development session. Between sessions, these teachers worked together in their schools to bring about changes, and reported these at the subsequent professional development sessions. Further work is needed to validate these observations and link the qualitative evidence with the quantitative data.

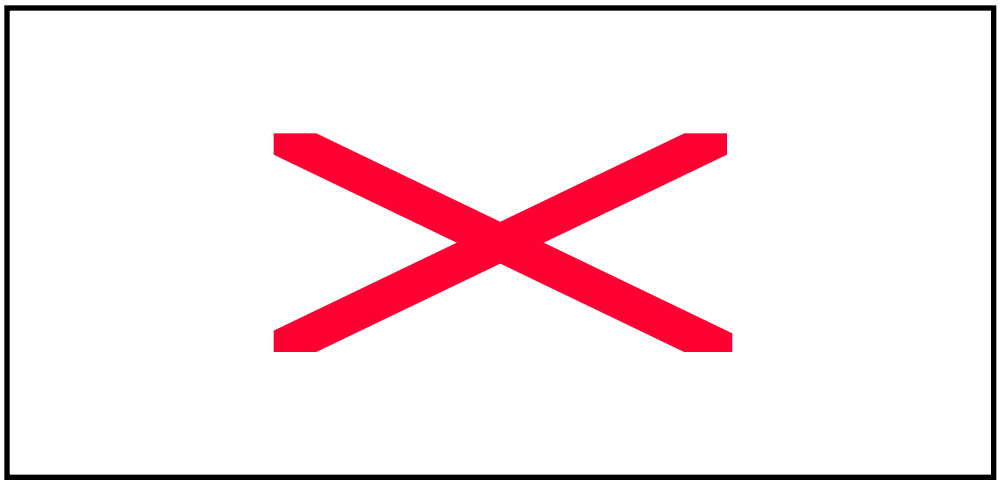

Figure 4. Mean infit by school.

Teacher Differences

Teachers who responded to the questionnaire were asked whether the INISSS program had changed their practice. Figure 5 shows these responses against the achievement of the students they were teaching, expressed as a mean logit score for the students.

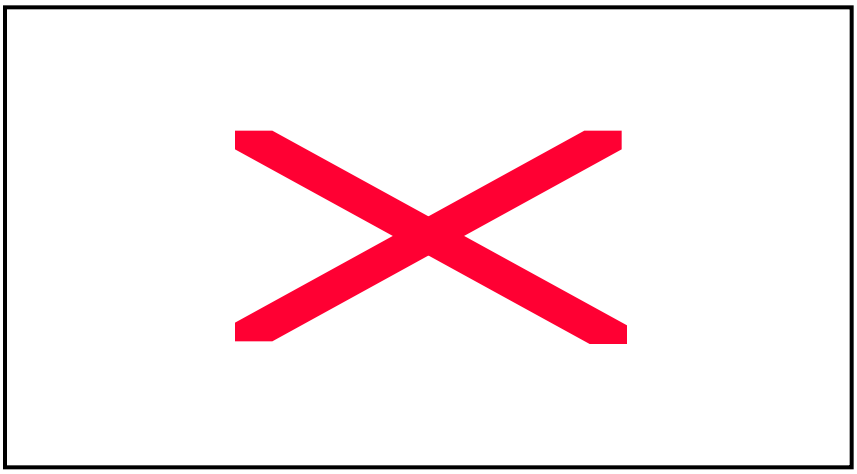

Figure 5: The effect of INISS PD on teachers' practice by student achievement

Those teachers who did not have an involvement with the INISSS program were consistently associated with the lowest student performance on the INISSS assessment tasks. The greatest change in performance was associated with those teachers who said that the INISSS professional development had changed their practice. There are several possible explanations. The INISSS program targeted teachers who were teaching the lowest levels of mathematics since the aim was to improve the performance of students in these groups. Many of these were teachers who had a limited mathematics background. It may be that participation in the INISSS program allowed those teachers to become more confident as teachers of mathematics so that by the A3 assessment they were teaching more able classes of students. There was some anecdotal evidence that this could be so.

An alternative explanation is that teachers who changed their practice became better at teaching the kinds of skills assessed by the INISSS assessment tasks – they "taught to the test". Since the INISSS assessment process mirrored the kinds of teaching approaches advocated in current curriculum documents this is of less concern than it might have been. In addition, since the questionnaire data were collected at the time of the A2 assessment, and no further professional development occurred in the interim between A2 and A3, it seems likely that students in Year 9 who were taught by teachers who had undertaken the INISSS program were experiencing the kinds of teaching approaches being advocated and that this was reflected in their performance in the A3 assessment. Given the student performance results, these kinds of approaches may benefit Indigenous girls in particular.

Discussion and Conclusion

In relation to the principal purpose of the evaluation, the INISSS program appeared to have met its goals of improving Indigenous students' numeracy outcomes. Although some improvement might have been expected over the 18 months between A1 and A3 due to natural growth, the improved performance of Indigenous students, and Indigenous girls in particular, goes against reported trends. One weakness of this evaluation study, however, is the lack of comparison with schools that were not involved, or some measure of improvement against some external standard. The original study design intended to link performance on the INISSS assessment tasks to that on statewide testing planned for Year

7 in 1998 and Year 9 in 2000. Although data are available from Year 7 testing for the target cohort, Year 8 in 1999, the planned statewide testing in Year 9 in 2000 did not eventuate. This highlights a difficulty in undertaking applied research within a system context where there are different priorities.

The assessment process, however, did address many of the concerns about performance assessment, the first subsidiary research question. All the assessment tasks measured the same construct in ways that were seamless with advocated teaching practice, addressing the validity issue. The Rasch model used showed little evidence of misfit, suggesting that the consistency issue had been overcome. The generalisability difficulty was attended to by placing results from all tasks on one scale. INISSS is one of the few professional development programs that has used students' outcomes as a major facet of its evaluation – providing a 'hard edge' to the evaluation process. As such, the measured improvement should be taken seriously since the assessment process was an integral part of the program, providing good information to teachers about their students' progress. Evidence from videotape of the professional development sessions suggests that teachers saw the assessment as important and changed their practice in ways that were beneficial to their students (Callingham 2000). However, although teachers were provided with information about their students' achievement after each assessment event, there was no attempt to link this with targeted teaching for maximum effectiveness. A more explicit connection between the outcomes from assessment and teaching practice, through embedding the assessment process into classroom practice, has the potential to further improve students' outcomes.

Finally, it appears that the INISSS professional development did change teachers' approaches to teaching. Where teachers reported changing their practice, this was associated with gains in students' achievement, answering the second subsidiary research question. It seems that undertaking the professional development did support improved outcomes for students. Students themselves recognised this. In videotaped discussion they identified changes in their mathematics classrooms, and expressed satisfaction with what was happening. As one student said, referring to a comment that maths was more enjoyable than it had been, "When it's fun, you learn".

This paper highlights some aspects of the evaluation of an extensive program that invested heavily in teachers by providing professional development and resources. One important aspect of the evaluation was the measurement of students' numeracy outcomes, using specially designed performance assessment tasks. Against these measures the INISSS program met its goals. The program changed some teachers' practice, and this was associated with gains in their students' performance. In addition it demonstrated that complex performance assessment tasks relying on teacher judgement can be successfully used for evaluation. The extent to which the assessment process itself contributed to the improvement in students' outcomes will be the subject of ongoing research.

References

Adams, R.J. & Khoo, S.T. (1996). *Quest: Interactive item analysis.* Melbourne: ACER.

Callingham, R. & Griffin, P. (2000). Towards a framework for numeracy assessment. In J. Bana & A. Chapman (Eds.), *Mathematics Education beyond 2000* (Proceedings of the 23rd Annual Conference of the Mathematics Education Research Group of Australasia, Fremantle, pp. 134-141). Sydney: MERGA.

Callingham, R. (2000). Inclusive practice in mathematics: Becoming comfortable with alternative assessment. In J. Wakefield (Ed.). *Mathematics: Shaping the future* (Proceedings of the 37th annual conference of the Mathematical Association of Victoria, pp. 125-135). Melbourne: MAV.

Callingham, R.A. (1999). Developing performance assessment tasks in mathematics: a case study. In J.M. Truran & K.M. Truran (Eds.) *Making the Difference.* (Proceedings of the 22nd annual conference of the Mathematics Education Research Group of Australasia. Adelaide, pp. 135-142). Sydney: MERGA.

Godfrey, J., Partington, G.,Haslett, M., Harrison, B. & Richer, K. (1998, December). *Attitudes of Aboriginal students to schooling*. Paper presented at the annual conference of the Australian Association for Research in Education, Adelaide.

Groome, H. & Hamilton, A. (1995). *Meeting the educational needs of Aboriginal adolescents* (Commissioned report No. 35). Canberra: NBEET.

Griffin, P. (1997). *Using the Rasch model*. Melbourne:Assessment Research Centre.

Hill, P.W., Rowe, K.J., Holmes-Smith, P. & Russell, V. (1996). *The Victorian Quality Schools Project. Report (Volume 1).* Melbourne: Centre for Applied Educational Research.

Hughes, P. & More, A.J. (1997, December). *Aboriginal ways of learning and learning styles.* Paper presented at the annual conference of the Australian Association Research in Education, Brisbane.

Lokan, J. (1999). Equity issues in testing: The case of TIMSS performance assessment. *Studies in Educational Evaluation*. 25, 297-314.

Linn, R.L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher, 23*(4), 4-14.

MCEETYA. (2000). *Achieving educational equality for Australia's Aboriginal and Torres Strait Islander peoples.* Canberra: Author.

McRae, D., Ainsworth, G., Cumming, J., Hughes, P., Mackay, T., Price, K., Rowland, M., Warhurst, J., Woods, D., & Zbar, V. (2000). *Education and training for Indigenous students: What has worked (and will again)*. Canberra: Australian Curriculum Studies Association and National Curriculum Services.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Moskovich, J.N. (1998). Rethinking authentic assessments of student mathematical activity. *Focus on Learning Problems in Mathematics, 20*, 1-18.

Nicholson, V. (1999, November). *The INISSS program*. Paper presented at the annual conference of the Australian Association for Research in Education, Melbourne.

Shepard, L.A. (2000), *The role of classroom assessment in teaching and learning* (CSE Technical Report 517). CRESST. Retrieved 18 March 2001 from http://www.cse.ucla.edu/CRESST/pages/reports.htm

Williams, D. (1997). *Mathematics task centre project for Aboriginal and Torres Strait Islander students*. Carlton: Curriculum Corporation.