Patterns of Performance and Progress: Analysis of 2004 data

Jenny Young-Loveridge
University of Waikato
<educ2233@waikato.ac.nz>

Data from approximately 70,000 year 1–8 students who participated in the Numeracy Development Project (NDP) in 2004 was analysed and compared with corresponding data from 2003 and earlier. As in previous years, all students seemed to benefit from participation in the project, but some groups made greater progress than others. Asian students were the highest performers, both in terms of percentages at the highest framework stages and in the progress they made relative to other students who began the project at identical starting points. Pākehā/European students did better than Māori students, who in turn outperformed Pasifika students. However, there was evidence to suggest that the gaps between groups may be getting smaller. As in previous years, students from high-decile schools did better than those from medium- and low-decile schools. However, students from low-decile schools who began the project at stages 0–3 made greater progress than comparable students from medium-decile schools. This may have been because certain low-decile schools were receiving additional support through another Ministry initiative. Analysis of data from adjacent year groups enabled the impact of the NDP to be separated from the effects of "normal" aging, and this showed younger students *after* the project to be significantly better than older students *before* the project. Effect sizes provided a measure of how practically meaningful the differences were. These were almost half a standard deviation for multiplication/division and proportion/ratio and about a quarter of a standard deviation for addition/subtraction.

Education systems worldwide have taken up the challenge to reform the teaching of mathematics in order to improve the mathematics learning of their students. The rhetoric that has accompanied such reforms has often justified them in terms of the need to produce citizens who are better able to cope with the demands of the twenty-first century (Bobis et al., 2005; British Columbia Ministry of Education, 2003; Commonwealth of Australia, 2000; Department for Education and Employment, 1999; National Council of Teachers of Mathematics, 2000; New South Wales Department of Education and Training, 2003; Ministry of Education, 2001). Recent evaluations of the reform process have highlighted the success of the reform efforts but drawn attention to some of the unintended consequences that indicate the need to modify approaches being taken (Earl et al., 2003).

New Zealand responded to the calls for reform in mathematics education by developing its Numeracy Development Project (NDP) approximately five years ago. This began initially with a small group of students in the early years of primary school (Early Numeracy Project [ENP] years 0–3), extended outwards to other schools, and then upwards into the senior primary years (Advanced Numeracy Project [ANP] years 4–6), the intermediate years (Intermediate Numeracy Project [INP] years 7–8), and most recently, the first two years of secondary school (Secondary Numeracy Project [SNP] years 9–10). By the end of 2005, it is expected that approximately 17 000 teachers and 460 000 students will have participated in the NDP (Parsons, 2005). It is predicted that by about 2007, virtually all teachers at years 0–6 and the majority of those at years 7–8 will have been given the opportunity to be involved in one of the professional development programmes as part of the NDP. Comprehensive evaluations have been undertaken of each of the professional development programmes that are part of the NDP (ENP: Thomas & Ward, 2002; Thomas, Tagg, & Ward, 2003; Thomas & Tagg, 2004; ANP: Higgins, 2002, 2003, 2004; INP & SNP: Irwin, 2003, 2004; Irwin & Niederer, 2002). An analysis of the overall patterns of performance and progress across years 0–8 was undertaken for the years 2001–2003 (Young-Loveridge, 2004). A Māori-medium version of the NDP was

also evaluated (Christensen, 2003, 2004). All of the evaluations have shown the NDP to be effective in raising mathematics achievement across primary and early secondary levels of the school system, and the benefits have been demonstrated in both Māori-medium and in English-medium settings.

A more fine-grained analysis was made possible by the aggregation of data across the three (English-medium) primary projects (Young-Loveridge, 2004). This analysis took students who began the NDP at the same stage on the Number Framework and looked at their progress over the course of the project as a function of gender, ethnicity, socio-economic status (as reflected in school-decile ranking), and year group. It became clear that not all groups benefited from the numeracy projects to the same extent. For example, Pākehā/European and Asian students made greater progress than Māori and Pasifika students, students at high-decile schools made greater progress than those at low- or medium-decile schools, boys tended to make greater progress than girls, and older students made greater progress than younger students. These findings showed that a "one size fits all" approach is not appropriate, and steps needed to be taken to tailor the projects to better meet the needs of particular groups of students. These findings were shared with all numeracy facilitators and consultants to reinforce the need for them to work with teachers in ways that helped schools shape the project to match the learning needs of their particular students.

This paper describes the findings from the analysis of data from year 0–8 students who participated in the project in 2004 and compares the patterns of performance and progress with those of students in 2003. The research question that guided this part of the project was:

> How does the performance and progress of students who participated in the numeracy projects in 2004 vary as a function of ethnicity, socio-economic status, and gender?

Method

Participants

Data from approximately 70 000 students who were assessed at the beginning and end of the NDP were included in the analysis. Just over one-third of the cohort was from ENP, almost half was from ANP, and the remaining students were from INP (see Appendix A). More than half of the students were Pākehā/European, about a fifth were Māori, a tenth were Pasifika, and the remainder were Asian or another ethnicity (see Appendix B). A third of the students were from high-decile schools, a quarter were from low-decile schools, and the remaining 40 percent were from medium-decile schools. The gender composition of the group was virtually identical. It was interesting to note that, compared to 2003, the 2004 cohort had slightly more Pākehā/European students and fewer Māori, as well as more students from medium- and high-decile schools and fewer from low-decile schools.

Procedure

Students were interviewed individually by their teachers at the beginning and end of the NDP using the diagnostic interview (NumPA), and the data was then sent to a secure website. Only students for whom there was complete data were included in the analysis for this report.

Findings

Patterns of Performance

The first part of the result examines students' performance, before and after the NDP and as a function of grouping variables such as age (reflected in year group), ethnicity, socio-economic status (reflected in school-decile band), and gender.

Differential performance as a function of year group

As in other years of the NDP, students tended to be assessed by their teachers as being at a higher framework stage after the project than they had been at the start (for details of the performance of each year group, see Appendix A). Performance improved steadily for each successive year group. By the end of the project, there was still substantial variation in performance across year groups. For example, on addition/subtraction, the percentage of students at the highest framework stage (stage 6, Advanced Additive Part–Whole) ranged from a fraction of a percent (0.1%) of year 1 students through to more than half of the year 8 students (55.1%). Some students, particularly the younger students, were not given the chance to show multiplicative thinking or proportional reasoning because Form A of the diagnostic interview (NumPA) was used, and the only operations assessed in Form A are addition and subtraction. By the end of the project, the proportion of students judged to be Advanced Multiplicative Part–Whole (stage 7) ranged from about four percent (3.7%) at year 4 through to a third of year 8 students (33.8%). The corresponding values for Advanced Proportional Part–Whole (stage 8) ranged from less than one percent (0.3%) at year 4 through to less than ten percent (9.3%) of year 8 students. The low levels of performance on the multiplicative and proportional domains have some important implications for the secondary schools that received these year 8 students into year 9 this year.

Differential performance as a function of ethnicity, decile, and gender

Appendix B shows the percentages of students at each framework stage on the various operational domains as a function of gender, ethnicity, and decile band. The most notable differences were evident at the highest framework stages. Consistently more boys than girls were at the highest framework stage, and this pattern held across all operational domains, both before and after the NDP. Out of the four main ethnic groups, Asian students performed the best, followed by European, then Māori, and finally Pasifika students. Again, the differences were very consistent across all domains and at both initial and final assessments. As a consequence, the relative differences among the various ethnic groups were maintained, and the gaps in performance between ethnic groups do not appear to have been narrowed appreciably by the project. When gender and ethnicity were examined together, the superiority of boys over girls was found consistently for all ethnic groups, though the magnitude of the gender difference varied somewhat from one group to another. Statistical analysis indicated that it was among European students that the gender difference was the greatest and most consistent. The tendency of Māori boys to outperform Māori girls in this project was contrary to the findings of many other projects. This suggests that individual diagnostic interviews, where both the presentation of tasks and students' responses to them are oral, may be a more valid assessment of the mathematical understanding of Māori boys than paper and pencil tests administered to large groups. In the past, Māori boys have done more poorly than Māori girls when mathematics is assessed using paper and pencil tests.

The variation in performance as a function of school-decile band was somewhat less consistent than the patterns for the other grouping variables. In previous years, performance

tended to increase with decile band. However, in 2004 the medium- and high-decile bands were very similar in their performance patterns in many instances.

Comparison with 2003 data showed that students did better in 2004 than they had done the year before. This could be explained by differences in the composition of the cohorts, with the 2004 cohort including more students from medium- and high-decile schools. Another possibility is that the NDP facilitators have become more effective as they gain more experience with the project. Anecdotal evidence from conversations with facilitators indicates that they perceive themselves to be more effective now than they were in 2001 when the NDP first began. It is possible that both a cohort effect *and* the increased effectiveness of facilitators have contributed to better performance in 2004.

The impact of the NDP on students' performance: analysis of effect sizes

It is not immediately clear from the analysis of performance whether students would have made progress simply as a result of "normal" aging rather than because they had been part of the NDP. In a traditional experimental design, comparisons are made of the progress (as measured by the difference between pre-test and post-test scores) for the "intervention" group that received the "treatment" and the "control" group that did not. When the NDP began, evaluators made a deliberate decision not to have a control group because of the ethics of withholding the programme from teachers and students who could benefit from it. Another reason for not having a control group was that the logistical problems of training non-participant teachers to assess their own students simply for the purpose of comparison with students whose teachers did participate in the project would have been great. An important dimension of the NDP is that the assessment of the students is done by their own teachers as part of the professional development programme that comprises the "intervention", and hence is an essential component of the intervention process itself. Getting outside researchers to interview a control group of students for comparison purposes would also have been problematic, as it would have introduced another potentially confounding variable to the comparison.

One way around the lack of a control group is to use data from the students before they began the project as a comparison with the data after they had finished the project. Gill Thomas, for example, used a "reference group" to compare the "growth in each aspect of number learning that occurred over the duration of the project with the growth that would have been expected with age alone" (Thomas & Ward, 2001, p. 14). Thomas found that the gains made during the project were greater than the gains "that would have been expected in the students' previous classroom programmes" (p. 14). A bar graph (Thomas & Ward, 2001, Figure 3.1) showed the overall difference between the gains of the project students and their reference group on addition/subtraction strategies. However, no test was done of the statistical significance of this difference nor any analysis of variation as a function of age or magnitude of effect size.

For the purposes of this report, an analysis was undertaken of adjacent year groups to explore the differences between younger students *after* the project and older students *before* the project. The professional development programme took place over about three school terms (approximately three-quarters of a calendar year), so by the end of the project, the younger students were, on average, about a quarter of a year younger than the older students with whom they were being compared. This meant that the students at the end of the project were still at a slight disadvantage developmentally, compared with their older peers before the project. Hence, any statistically significant differences in favour of younger students after the project should reflect real and notable benefits to these students as a result of participating in the project.

Before presenting the comparison of younger students after with older students before, a simple comparison was done of each year group before the project to ascertain the pattern of

"normal" development without intervention. This data provides a baseline for the other comparisons. Figure 1 shows the average framework stage for each year group before the project. In the early school years, there was a difference of about one stage between each adjacent year group. This decreased with age to about a fifth of a stage by intermediate because of a ceiling effect operating for addition/subtraction. In spite of a reduction in magnitude, all differences between adjacent year groups were statistically significant at or beyond the 0.001 level (see Appendix C). Effect sizes (based on the standardised mean difference between groups) are reported here in Table 1 and provide a measure of how practically meaningful the differences were, because it is well known that a large sample can yield statistically significant results that are not practically meaningful (see Fan, 2001). According to Fan, an effect size of 0.20 is "small", 0.50 is "medium", while 0.80 is "large". Table 1 shows effect sizes that were large initially (–0.83), but diminished to medium by about year 4 (–0.58 to –0.34), and small (–0.20) by year 8. It is likely that ceiling effects helped to reduce the magnitude of effect sizes in the senior primary and intermediate years.

Table 1

Average Framework Stages and Corresponding Effect Sizes for Younger and Older Students in Adjacent Year Groups (2004)

Addition/Subtraction						Multiplication/Division			Proportion/Ratio		
Year Groups	Younger **Before**	Older **Before**	**Effect Size**	Younger **After**	**Effect Size**	Younger **After**	Older **Before**	**Effect Size**	Younger **After**	Older **Before**	**Effect Size**
1 & 2	1.52	2.48	**–0.83**	2.54	**0.06**						
2 & 3	2.48	3.44	**–0.76**	3.50	**0.05**	4.18	3.73	**0.59**	4.17	3.75	**0.70**
3 & 4	3.44	4.13	**–0.58**	4.24	**0.10**	4.45	4.14	**0.36**	4.37	4.06	**0.42**
4 & 5	4.13	4.48	**–0.34**	4.69	**0.23**	4.83	4.51	**0.33**	4.69	4.34	**0.40**
5 & 6	4.48	4.69	**–0.22**	4.95	**0.30**	5.21	4.84	**0.35**	5.02	4.65	**0.38**
6 & 7	4.69	4.85	**–0.16**	5.17	**0.38**	5.57	5.12	**0.43**	5.40	4.95	**0.40**
7 & 8	4.85	5.03	**–0.20**	5.23	**0.23**	5.71	5.39	**0.31**	5.58	5.25	**0.28**
Average			***–0.44***		***0.19***			***0.40***			***0.43***

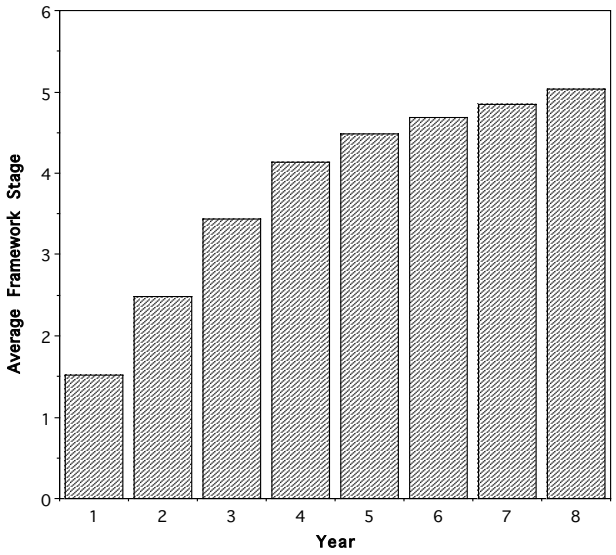

Figure 1. Average framework stage on *Addition/Subtraction* for each year group *before* participating in the project

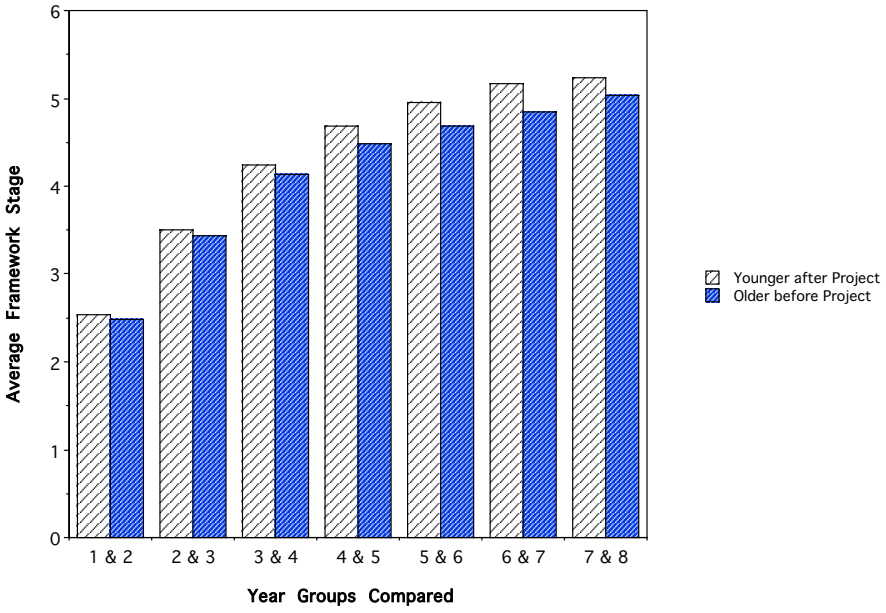

Figure 2. Average framework stage on *Addition/Subtraction* for younger students *after* the project compared with older students *before* the project

Figure 2 shows the average framework stages on addition/subtraction of adjacent year-groups for the younger students *after* the project compared with older students *before* the project (the 3rd and 5th columns of Table 1 show the framework stages used in the comparison, while the 6th column shows the effect sizes). All of the comparisons were statistically significant at the 0.001 level (see Appendix C). It is interesting to note that the magnitude of the effect size increased steadily with each year group up to a maximum of approximately a third of a standard deviation at year 6 and 7, then declined slightly at year 7 and 8, again possibly because students in year 7 and 8 are generally at the higher levels of the framework, producing a ceiling effect. This pattern of increase is consistent with the idea that older students benefit more from the project than younger students, despite anecdotal evidence suggesting that the project is easier to implement at junior primary levels than at the senior end of the school. The effect sizes for addition/subtraction ranged from very small (0.05) to moderate (0.38), with an average of 0.19. Part of the reason for smaller effect sizes at younger ages may have been the greater variability within the groups being compared because progression at lower framework stages is easier but may be less reliably assessed. At older age groups, there is a tendency to be at higher framework stages and progression to a higher stage is harder but more clear-cut and hence more reliably assessed.

A similar analysis was done for multiplication/division and proportion/ratio (see the 7th to 12th columns in Table 1). Figures 3 and 4 present the average framework stages on multiplication/division and proportion/ratio of adjacent year-groups from years 2 to 8 for 2004 (for details, see Appendix C). The effect sizes for multiplication/division and proportion/ratio were mostly within the moderate to fairly large range (0.28 to 0.70), with averages of 0.40 and 0.43 for multiplication/division and proportion/ratio, respectively. It was interesting to note that there was a particularly large difference at years 2 and 3 (larger than at years 3 and 4), then a steady increase up to years 6 and 7, followed by a slight decline. The large difference initially may have been the result of such a small proportion of year 2 students having been assessed on multiplication/division and proportion/ratio (approximately 16% of the year group), and the fact that these students must have been extremely good mathematicians for their age. This can be concluded from the fact that students must have impressed their teachers sufficiently to be assessed in domains not usually taught at their year level (using form B or form C of the diagnostic interview). As Appendix A shows, initially only 16 percent of year 2 and 53 percent of year 3 were given the chance to show multiplicative strategies or proportional reasoning. By year 4, virtually the whole year group was assessed on all three domains. Hence, the figures used to calculate effect size include students from the full range of mathematical abilities. It was interesting to note also that the benefits for students on multiplication/division and proportion/ratio were greater than on addition/subtraction, both in terms of the difference in average framework stage and effect size. The reason for the greater effect sizes may have been that there was much less of a ceiling effect operating for these domains than was evident for addition/subtraction.

Patterns of Progress

Patterns of progress were examined by looking at the proportion of students that moved up to a higher framework stage relative to a particular starting point. This analysis was done for ethnicity, decile band, and gender (see Appendix C).

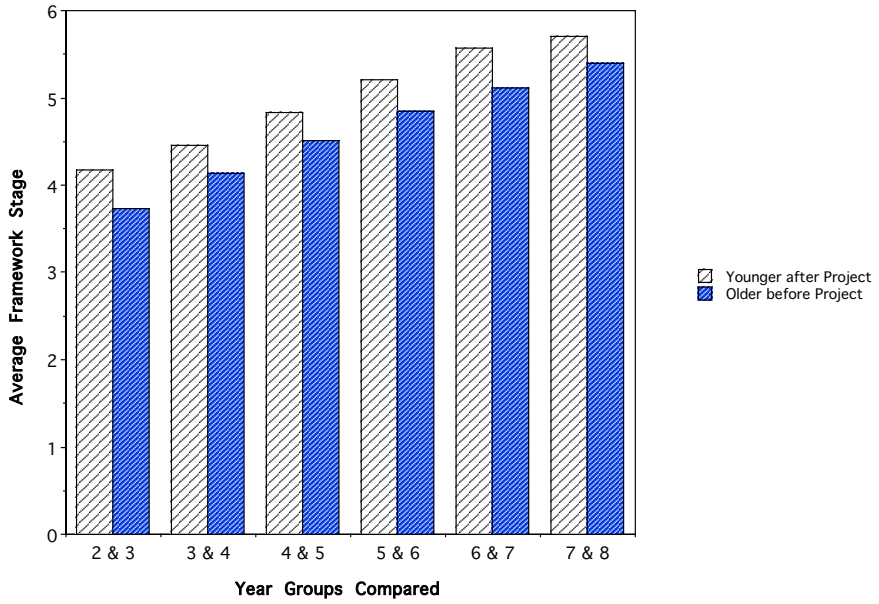

Figure 3. Average framework stage on *Multiplication/Division* for adjacent year groups

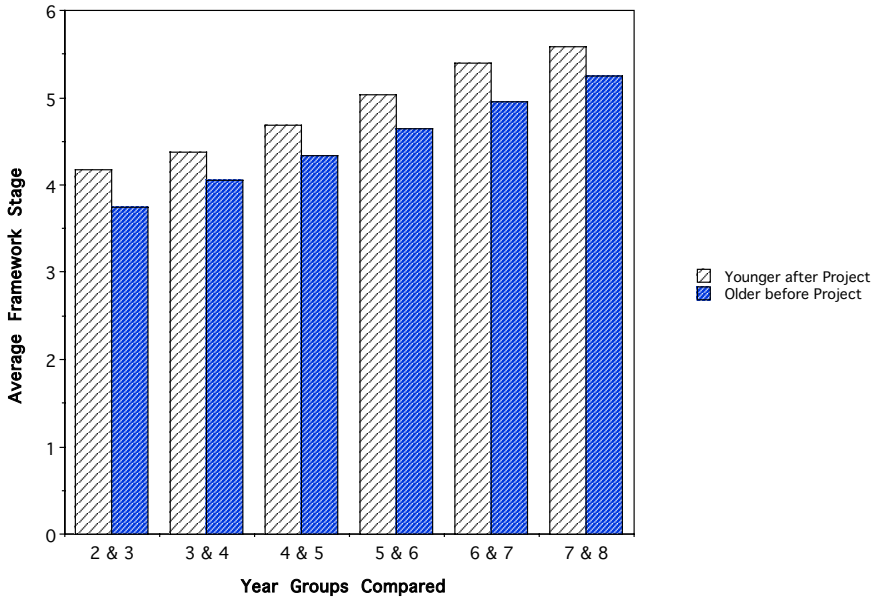

Figure 4. Average framework stage on *Proportion/Ratio* for adjacent year groups

The impact of ethnicity on progress

Figures 5 and 6 show the patterns of progress on addition/subtraction as a function of ethnicity. Asian students consistently made the greatest gains, followed by Pākehā/Europeans, with Māori and Pasifika students gaining the least. Progress to a higher stage was greater in 2004 for all ethnic groups than it had been in 2003, with the smallest increase for Europeans (1.9%) and the greatest increase for Pasifika (6.7%) (see Figure 6). The result was a narrowing of the gap between European and Māori students (6.1% to 2.9%) and between European and Pasifika students (8.4% to 3.6%). The gap between Māori and Pasifika students also narrowed (2.3% to 0.7%). Analysis of effect size was also undertaken (see Table 2 below). It was interesting to note that, although there were statistically significant differences in progress between ethnic groups, the magnitude of the effect size for comparisons of European students with Māori, and European students with Pasifika, was relatively modest (average = 0.13 and 0.17, respectively). The difference between the most successful group (Asian) and the least successful group (Pasifika) was, on average, just over a third of a standard deviation (0.36). The effect sizes for corresponding comparisons done in 2003 were identical for European with Māori, larger for European with Pasifika (0.25 vs 0.17), but smaller for Asian with Pasifika (0.31 vs 0.36) (see Table 2). This could be explained by an improvement in the progress of Pasifika students, accompanied by an even greater improvement in the progress of Asian students.

The impact of school decile on progress

Figure 7 (see page 15) shows the patterns of progress on addition/subtraction as a function of school decile. The most striking finding for this analysis is that students at low-decile schools who began the project at stage 3 or lower made greater progress than those from medium-decile schools (44.2% of low-decile students went up at least a stage compared to 39.9% of medium-decile students) and almost as much progress as students at high-decile schools (44.7%; see Appendix B, Table B7). This is very different from the pattern in 2003 when students at low-decile schools consistently made the least progress while those at high-decile schools made the most (see Figure 8 and Young-Loveridge, 2004). For students who began the project either counting on or using simple partitioning strategies, the pattern was more similar to the previous year, with students at high-decile schools making the most progress and those at low-decile schools the least. Table 2 shows effect sizes for the difference between the high- and low-decile bands in 2004 and 2003. It is clear from Table 2 that the difference halved from 2003 to 2004 (0.22 to 0.10, on average).

Table 2

Effect Sizes for Differences between Subgroups on Addition/Subtraction (2004 & 2003)

2004							2003			
Ethnicity				Decile	MEI*		Ethnicity			Decile
Initial Stage	Eur vs Māori	Eur vs Pasifika	Asian vs Pasifika	High vs Low	MEI vs nonM		Eur vs Māori	Eur vs Pasifika	Asian vs Pasifika	High vs Low
0	0.31	0.42	0.86	-0.44	-0.52		0.15	0.09	0.53	0.27
1	0.10	0.01	0.21	0.11	0.57		0.08	0.10	0.26	0.17
2	0.03	0.07	0.31	0.12	-0.06		0.13	0.17	0.21	0.18
3	0.10	0.13	0.20	0.13	0.24		0.13	0.25	0.25	0.26
4	0.09	0.19	0.25	0.17	0.18		0.12	0.38	0.31	0.23
5	0.14	0.19	0.34	0.15	0.14		0.14	0.49	0.31	0.21
Average	***0.13***	***0.17***	***0.36***	***0.10***	***0.05***		***0.13***	***0.25***	***0.31***	***0.22***

* Manurewa Enhancement Initiative

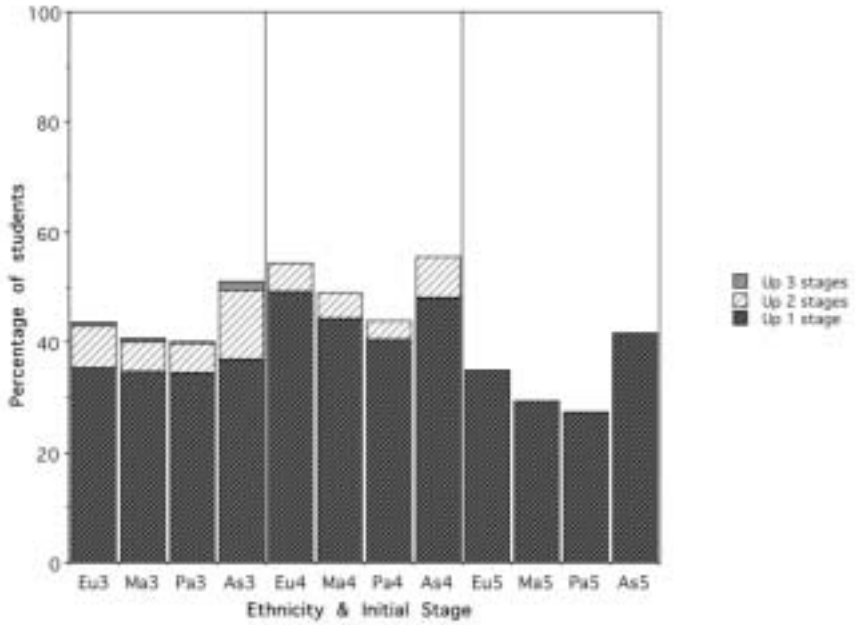

Figure 5. Percentage of students who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial stage and *ethnicity* (2004)

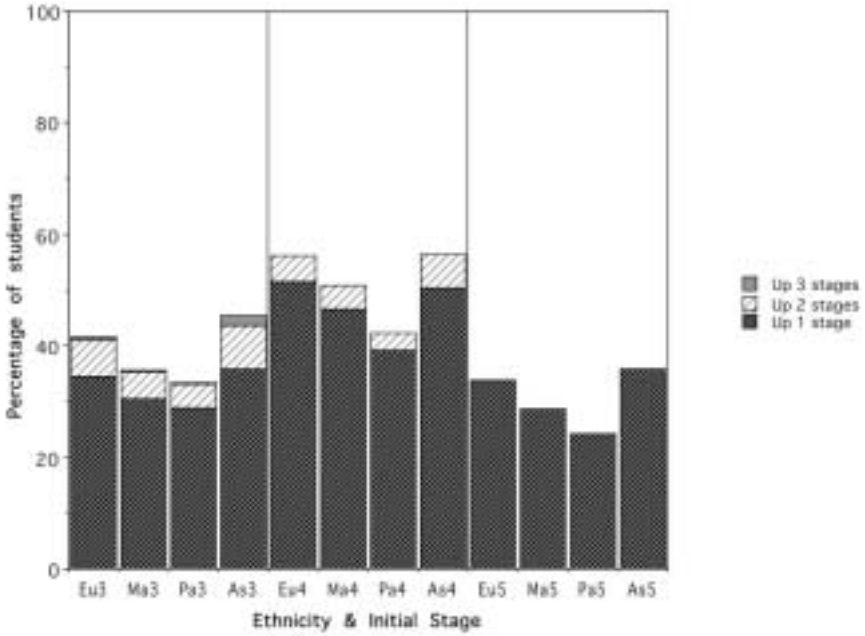

Figure 6. Percentage of students who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial stage and *ethnicity* (2003)

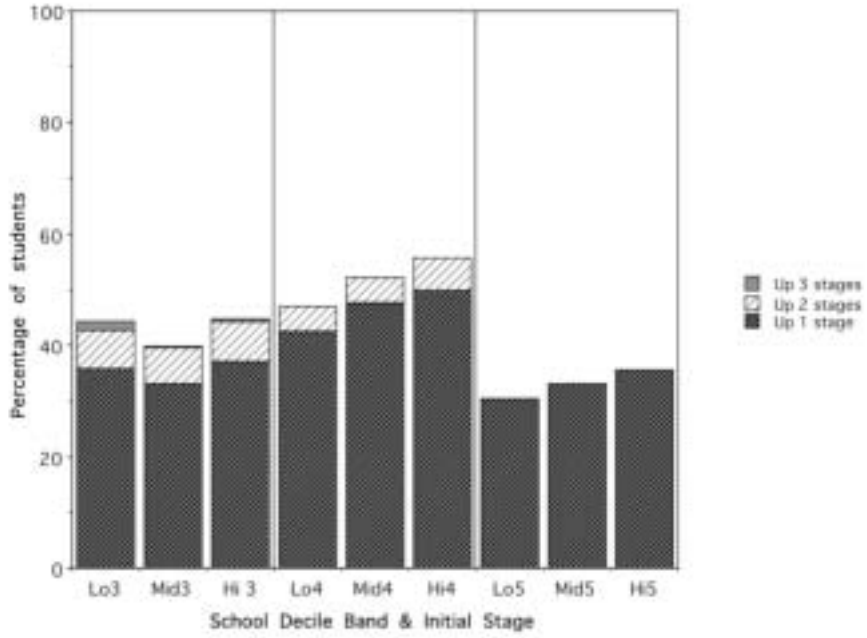

Figure 7. Percentage of students who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial stage and *decile* band (2004)

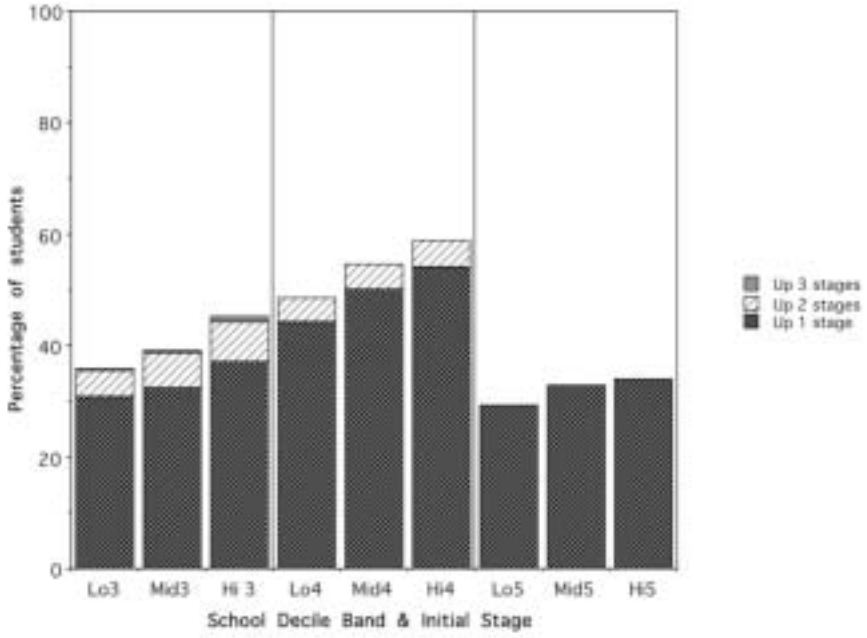

Figure 8. Percentage of students who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial stage and *decile* band (2003)

The possible impact of other initiatives on some low-decile schools

Other research has shown that students at low-decile schools have lower levels of achievement than those at medium- or high-decile schools (see Alton-Lee, 2003). The "decile" system that assigns a ranking of 1 to 10 to a school on the basis of census information about the income and educational levels in the mesh blocks in which its students reside was developed to enable more funding to be provided to more disadvantaged schools. Over the last few years, there have been various School Improvement initiatives operating quite independently of the NDP. An analysis done for this report indicates that some of the improvements for students at low-decile schools could have been the result of one of the special initiatives that was put in place in 2004 to provide extra support for schools in certain low-income areas. The Manurewa Enhancement Initiative (MEI) was a schooling improvement initiative focusing on integration and alignment with the NDP and had as one of its goals "added value", rather than just implementing the NDP in the normal way. Eight low-decile primary schools with complete data were identified from the list of MEI schools for this analysis. The patterns of progress for students at the eight MEI schools ($n = 942$) were compared with the corresponding patterns for the students at other low-decile schools ($n = 17\ 329$; see Table C5 in Appendix C). The patterns of progress were somewhat inconsistent and seemed to depend on students' starting points. For example, MEI students who began the project at stage 1 (One-to-One Counting) made greater progress than that made by other students at low-decile schools who also started at this stage [$t(58) = 3.62$, $p < 0.01$]. Those who started at stage 3 (Counting from One) also made significantly greater progress than that made by other comparable students [$t(67) = 2.08$, $p < .05$]. A similar pattern was evident for students who began the project at stage 4 (Counting On) [$t(437) = 3.64$, $p < 0.001$]. The opposite pattern was found at other starting points, including stage 5, Early Additive Part–Whole. Hence the effect sizes ranged from –0.52 to 0.57, averaging out at 0.05. The reason for the inconsistencies could be that the number of MEI students who began the project at stage 5 was relatively small compared with those in low-decile schools nationally. It may also be that the focus was more on developing increasingly efficient counting strategies and providing experiences with partitioning and recombining small quantities rather than working with multi-digit quantities.

The impact of gender on progress

Figure 10 shows the patterns of progress on addition/subtraction as a function of gender and initial stage. As in 2003, there appeared to be a small difference between boys and girls who started the project at a Counting All stage (stage 3) or below (favouring girls), whereas for those starting at stages 4 or 5 (counting on or simple partitioning strategies), boys made significantly greater gains than girls. Analysis of effect sizes for each starting point on each operational domain shows that the only significant gender difference that favoured girls was for those who were initially Emergent on addition/subtraction (stage 0). The average effect size on addition/subtraction was 0.05.

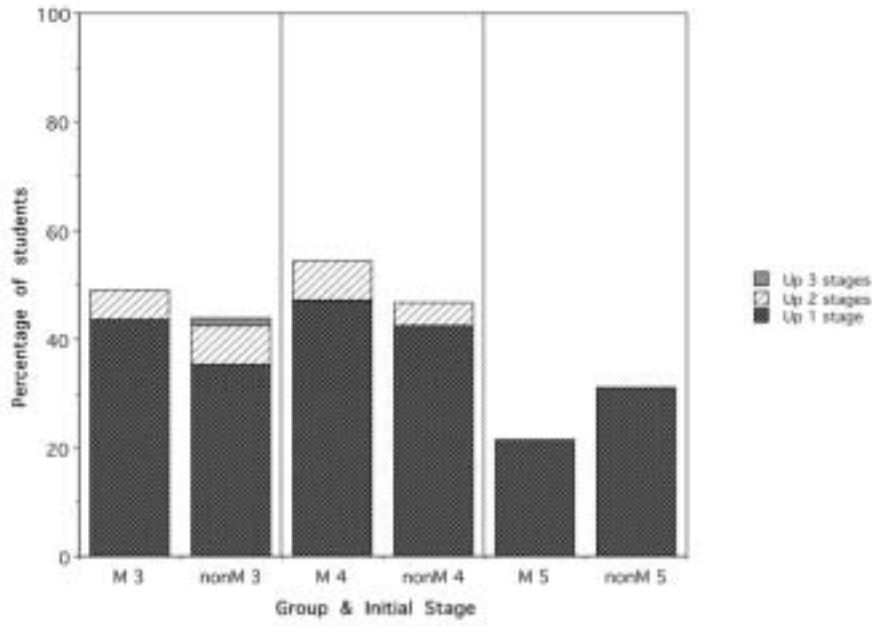

Figure 9. Percentage of students at low-decile schools who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial stage and *group* (schools involved in the Manurewa Enhancement Initiative vs others)

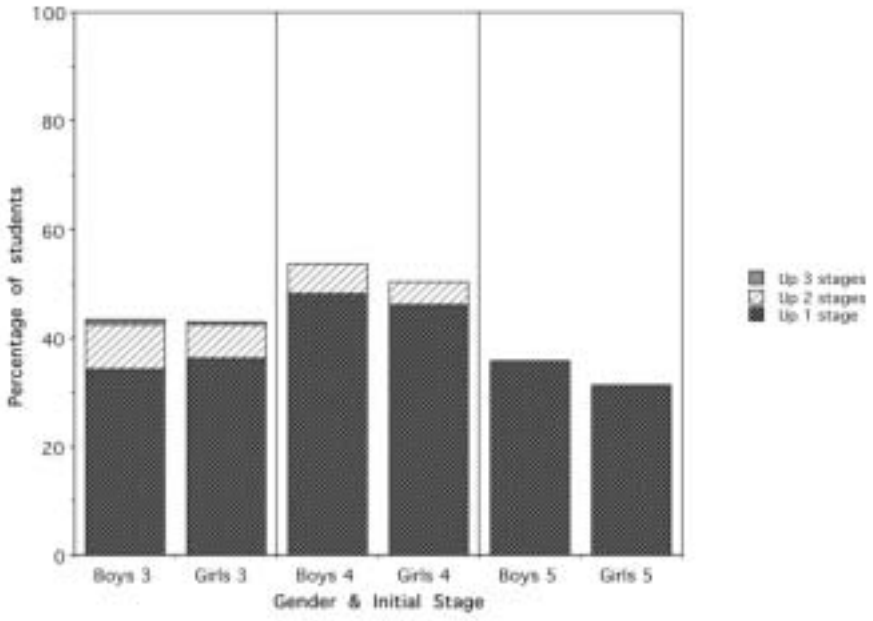

Figure 10. Percentages of students who progressed to a higher framework stage on *Addition/Subtraction* as a function of initial framework stage and *gender*

On multiplication/division, boys made significantly better progress than girls at each of the starting points, and the average effect size was 0.09. On proportion/ratio, boys again made significantly better progress than girls for all starting points except stage 7 (Advanced Multiplicative Part–Whole). The average effect size for proportion/ratio was 0.08.

Discussion

The analysis reported in this paper has shown that students who participated in the NDP in 2004 made significantly better progress on the Number Framework than would have been expected simply as a result of getting older. The advantage of being involved in the NDP was so great that it put younger students significantly ahead of slightly older peers who had not yet participated in the project. The calculation of effect sizes allowed the magnitude of differences to be examined for various different outcome measures and subgroups. The average effect size for addition/subtraction was 0.19, a relatively modest value, but very similar to that found for the National Numeracy Strategy in the UK (0.17 or 0.18; see Brown et al., 2003). However, average effect sizes for multiplication/division and proportion/ratio were more than double (0.40 & 0.43, respectively). It should be remembered that the effect sizes in the present study have been calculated using as a control group students who were, on average, a quarter of a year older than the students in the "experimental" group, and hence the effect sizes are very conservative measures of the impact of the "treatment" on students' performance.

There is some evidence in the 2004 data that shifts are beginning to occur in the patterns of progress found for some groups of students. Students from low-decile schools who began the project at stage 3 (Counting from One) or lower made significantly greater progress than students from medium-decile schools who also began the project at stage 3 or lower. Māori and Pasifika students made slightly better progress in 2004 than in 2003. These patterns (for low-decile and Māori/Pasifika students who began the project at or below stage 3) could be explained by the fact that additional support and resources were provided for low-decile schools in certain regions with a high concentration of low-decile schools (Ministry of Education, personal communication). A similar pattern was evident among students who began the project at the Counting On stage, although the progress of low-decile students was no greater than for medium-decile schools. The analysis of data for students involved in the Manurewa Enhancement Initiative shows how additional support and resources can make an even greater difference for students at low-decile schools. This is similar to the findings of a study that examined the impact of a major literacy initiative that succeeded in raising teachers' expectations of students' achievement in the early school years, and improved their students' literacy skills (*Picking Up the Pace*, see McNaughton et al., 2000, 2003; Phillips et al., 2002). Teachers in these schools have been able to see for themselves that it is indeed possible to change the educational outcomes for students from economically disadvantaged backgrounds, providing close attention is paid to meeting the students' particular learning needs in the classroom.

One of the strengths of the NDP is that it has evolved in response to feedback from the project evaluators and facilitators. Anecdotal evidence suggests that as the numeracy facilitators have gained more experience and understanding about the project, they have become increasingly effective in their work with teachers. At the beginning of the project, lower decile schools were given priority for inclusion in the project. However, it is the schools that have participated in the project in the more recent years that have benefited most from the accumulated wisdom of the facilitators. Ironically, more recent cohorts include disproportionately more high-decile schools and fewer low-decile schools. Teachers in low-decile schools, more often than those in medium- or high-decile schools, have additional issues to deal with on top of meeting the learning needs of their students in classrooms (for further details, see Ritchie, 2004). Although there have been efforts to provide further support for low-decile schools that participated in the project in earlier years (i.e., "sustainability" funding), it is not clear that these efforts have been sufficient to maintain the original impetus of the project. As the literature on educational reform shows, changing the ways that things are done in classrooms and schools is an extremely difficult and challenging process. However, it is to be hoped that the shifts beginning to occur in patterns of progress for students from lower decile schools will be sustained in subsequent years.

References

Alton-Lee, A. (2003). *Quality teaching for diverse students in schooling: Best evidence synthesis*. Wellington: Ministry of Education.

Bobis, J., Clarke, B., Clarke, D., Thomas, G., Wright, R., Young-Loveridge, J., & Gould, P. (2005). Supporting teachers in the development of young students's mathematical thinking: Three large scale cases. *Mathematics Education Research Journal*, *16* (3), 27–57.

British Columbia Ministry of Education (2003). *Supporting Early Numeracy: BC Early Numeracy Project (K-1)*. Province of British Columbia: Author.

Brown, M., Askew, M., Millett, A., & Rhodes, V. (2003). The key role of educational research in the development and evaluation of the National Numeracy Strategy. *British Education Research Journal*, *29* (5), 655–672.

Christensen, I. (2003). *An evaluation of Te Poutama Tau 2002*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Christensen, I. (2004). *An evaluation of Te Poutama Tau 2003*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Commonwealth of Australia (2000). *Numeracy, a priority for all: Challenges for Australian schools: Commonwealth numeracy policies for Australian schools*. Commonwealth of Australia.

Department for Education and Employment (1999). *The National Numeracy Strategy: Framework for teaching mathematics from reception to year 6*. London: Department for Education and Employment.

Earl, L., Watson, N., Levin, B., Leithwood, K., Fullan, M., & Torrance, N. (2003). *Watching and learning 3: Final report of the external evaluation of England's National Literacy and Numeracy Strategies*. Accessed on 27 March 2005 at: http://www.dfes.gov.uk/research/programmeofresearch

Fan, X. (2001). Statistical significance and effect size in educational research: Two sides of a coin. *Journal of Educational Research*, *94* (5), 275–282.

Higgins, J. (2002). *An evaluation of the Advanced Numeracy Project 2001*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Higgins, J. (2003). *An evaluation of the Advanced Numeracy Project 2002*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Higgins, J. (2004). *An evaluation of the Advanced Numeracy Project 2003*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Irwin, K. C. & Niederer, K. (2002). *An evaluation of the Numeracy Exploratory Study (NEST) and the associated Numeracy Exploratory Study Assessment (NESTA) years 7–10, 2001*. Wellington: Ministry of Education.

Irwin, K. C. (2003). *An evaluation of the Numeracy Project for years 7–10, 2002*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Irwin, K. C. (2004). *An evaluation of the Numeracy Projects for years 7–9, 2003*: Exploring issues in mathematics education. Wellington: Ministry of Education.

McNaughton, S., Phillips, G., & MacDonald, S. (2000). Curriculum channels and literacy development over the first year of instruction, *New Zealand Journal of Educational Studies*, 35, 49–59.

McNaughton, S., Phillips, G., MacDonald, S. (2003). Profiling teaching and learning needs in beginning literacy instruction: The case of students in "low-decile" schools in New Zealand, *Journal of Literacy Research*, 35, 703–730.

Ministry of Education (2001). *Curriculum update no. 45: The numeracy story*. Wellington: Learning Media.

New South Wales Department of Education and Training (2003). *Count Me In Too: Professional development package*. Sydney: New South Wales Department of Education and Training.

Parsons, R. (2005). *Numeracy Development Project: Scope and scale.* Paper presented at the Numeracy Development Project Reference Group, 15 March, Wellington.

Phillips, G., McNaughton, S., & MacDonald, S. (2002). *Picking up the pace: Effective literacy interventions for accelerated progress over the transition into decile 1 schools*. Wellington: Ministry of Education.

Ritchie, G. (2004). Teacher movement between schools in New Zealand: To high SES schools that have, more shall be given. *Journal of Educational Policy*, *19* (1), 57–79.

Thomas, G., & Tagg, A. (2004). *An evaluation of the Early Numeracy Project 2003:* Exploring issues in mathematics education. Wellington: Ministry of Education.

Thomas, G., Tagg, A., & Ward, J. (2002). Making a difference: The Early Numeracy Project. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds), *Mathematics Education in the South Pacific* (Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland, pp. 49–57).

Thomas, G., Tagg, A., & Ward, J. (2003). *An Evaluation of the Early Numeracy Project 2002*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Thomas, G. & Ward, J. (2001). *An Evaluation of the Count Me In Too Pilot Project*. Wellington: Ministry of Education.

Thomas, G. & Ward, J. (2002). *An Evaluation of Early Numeracy Project 2001*: Exploring issues in mathematics education. Wellington: Ministry of Education.

Young-Loveridge, J. (2004). *Patterns of performance and progress on the Numeracy Projects 2001–2003: Further analysis of the Numeracy Project data*. Exploring issues in mathematics education. Wellington: Ministry of Education.