Patterns of Performance and Progress on the Numeracy Development Project: Looking Back from 2005

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

This paper reports on the analysis of 2005 data from the Numeracy Development Projects (NDP). Students' performance improved from the beginning of the year to the end. However, performance on the multiplication and division domain was a little disappointing, with only about a third of year 8 students reaching stage 7, advanced multiplicative thinking, by the end of the year. Likewise, performance on proportion and ratio was lower than hoped, with only about a tenth of year 8 students reaching stage 8, advanced proportional reasoning. Analysis of the patterns of performance and progress over time showed that performance improved from 2002 to 2005. Students also made greater progress relative to their initial stage in recent years compared with earlier years. The gaps between European and Pasifika students appeared to reduce fairly steadily over time. Among students who began the project at lower Framework stages, Pasifika students often made greater progress than those from other ethnic groups. These improvements coincided with changes in the composition of the cohort over time, most notably a reduction in the percentage of students from low-decile schools and an increase in the percentage of students from medium- and high-decile schools. Hence, it is difficult to conclude with any confidence that the Numeracy Projects are primarily responsible for the improvements. Although the gaps in achievement between European and Māori/Pasifika students continued to exist, when these differences are put beside those found in other large-scale studies, it is clear that NDP differences are much smaller (a quarter of a standard deviation versus three-quarters to a whole standard deviation). The use of an individual, orally presented assessment tool with an emphasis on explaining the strategies used to get answers, rather than a written test on which the number of correct answers is simply totalled, may help to explain the positive outcomes for NDP students.

Background

The New Zealand Numeracy Development Projects (NDP), like other reforms in mathematics education world worldwide (for example, 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, 2001), came about as a result of concern about the quality of mathematics teaching. This concern was sparked by the results from the Third International Mathematics and Science Study (TIMSS), which showed that the mathematics achievement of students in many western nations was below international averages (for New Zealand's results, see Garden, 1996, 1997). Much of the rhetoric is about the need to produce confident life-long learners who are better able to cope with the demands of the twenty-first century.

The first Numeracy Development Projects began approximately six years ago. They are part of a wave of educational reform worldwide aimed at improving the mathematics teaching and learning of students at the primary and secondary levels. This paper reports on the results for the NDP for 2005.

Method

Participants

Data from approximately 52 000 year 0–8 students who were assessed at the beginning and end of the year 2005 were included in this analysis (see Appendix A for the composition of the sample).

Just under a third (29.9%) of the cohort were in years 0–3 (Early Numeracy Project: ENP), a little under half (46.4%) were in years 4–6 (Advanced Numeracy Project: ANP), and almost a quarter (23.7%) were in years 7–8 (Intermediate Numeracy Project: INP). More than half of the students (63.3%) were Pakeha/European, and about a fifth (19.5%) were Māori. The remainder consisted of 7.3% Pasifika, 5.4% Asian, and 4.4% of other ethnicities. Only about one sixth (18.1%) of the students were from low-decile schools, nearly half (44.7%) were from medium-decile schools, and more than a third (37.2%) were from high-decile schools. The gender composition was virtually identical. It was interesting to note that, compared to previous years, there were more European students and fewer Pasifika students. In 2005, there were fewer Māori students than in the initial years of the project (2002 and 2003). Other changes in the composition of the cohort included fewer students from low-decile schools and more from high-decile schools, an increase in the average decile ranking for all decile bands – low, medium, and high – and fewer students at years 0–3 and more at years 4–6 and 7–8.

Procedure

Students were interviewed individually by their teachers at the beginning and end of the year, using the diagnostic interview (Numeracy Project Assessment: NumPA). The data was sent to a secure website. Only students with two sets of data (initial and final) were included in the analysis for this report.

Results & Discussion

Patterns of Performance

The first part of this paper examines students' performance at the beginning and end of the year 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

At the end of the year, many students were at a higher stage of the Number Framework than they had been at the start of the project (see Appendix B). Comparison of 2005 results with those from 2004 show similar percentages of students reached equivalent stages on the Framework (see Young-Loveridge, 2005). For example, on the addition and subtraction domain, just over 80% of year 2 students in 2005 were at stage 3, counting from one by imaging, or higher, compared with 79% of year 2 students in 2004. Just over 95% of year 4 students in 2005 were at stage 4, advanced counting, or higher, and more than 65% were at stage 5, early additive part–whole thinking, or higher, compared with 95.5% and 63% of year 4 students in 2004 at stage 4 and stage 5 respectively. Almost 88% of year 6 students were at stage 5, early additive part–whole thinking, or higher, compared with almost 83% in 2004. Almost 59% of year 8 students in 2005 were at stage 6, advanced additive part–whole, compared with just over 55% in 2004. It is important to note here that most of the students were in their first year of the project and that even better results may be expected once the NDP has been established in a school over several years (see Thomas and Tagg, p. 22 in this compendium, on the results of the Longitudinal Study).

On the multiplication and division domain, the percentage of year 8 students at stage 7, advanced multiplicative part–whole thinking, at the end of the year was 36%, similar to the results found for students in 2004 (34%). Combining stages 7 and 8 on the proportion and ratio domain yielded almost identical percentages (36% in 2005 and 33% in 2004). These results are very consistent in showing that only about one-third of the year 8 students reached stage 7 by the end of the year.

The newly introduced stage 7 in the addition and subtraction domain (referred to as stage 7 A/S to distinguish it from stage 7 on other domains) involved addition and subtraction with fractions and decimals ($2 - [\frac{3}{4} + \frac{7}{8}]$ and $5.3 - 2.89$). These tasks seemed considerably harder than the multiplication and division tasks used to assess multiplicative thinking in the multiplication and division and proportion and ratio domains. Only 5% of year 8 students were initially at stage 7 A/S, and by the end of the year this had increased to only 19%. Comparable figures for stage 7 M/D were about 15% and 35% for initial and final data respectively. This discrepancy suggests that addition and subtraction with fractional numbers, particularly when the task involves decimals with place-holding zeros, is more difficult than multiplication and division with whole numbers and may be between stage 7 and stage 8 (see later section on inter-relationships between multiplicative thinking and other domains). The fact that more students were successful on addition and subtraction with fractional number than reached stage 8, advanced proportional part–whole thinking, indicates that it is easier to add and subtract with fractional numbers than to work with proportion and ratio.

The newly introduced stage 8, advanced proportional part–whole, in the multiplication and division domain (involving division with decimals) seemed to be of comparable difficulty to the proportion and ratio tasks used to assess proportional reasoning in the proportion and ratio domain. Only 2.2% of year 8 students were successful on these tasks initially, but this had increased to 10.8% by the end of the year. The corresponding figures for the proportion and ratio domain were 2.4% (initially) and 8.4% (finally). These figures were similar to those found in 2004 (2.8% initially and 9.3% finally). Hence by the end of the year, fewer than 10% of the students about to enter secondary schools were at stage 8.

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

Appendix C shows the percentages of students at each Framework stage for each domain as a function of gender, ethnicity, and decile. As in previous years, more boys than girls reached the upper stages of the Framework. For example, on addition and subtraction, the percentages of boys reaching stage 6 or higher were 14.7% initially and 30.5% finally, compared with 8.4% of girls initially and 22.4% finally. Similar gender differences were evident on the other domains as well. On multiplication and division, 6.7% of boys reached stage 7 or higher initially, compared with 2.9% of girls. The corresponding percentages at the end of the year were 16.9% and 10.9%. On proportion and ratio, 0.8% of boys reached stage 8, compared with 0.3% of girls. The corresponding percentages at the end of the year were 15.8% and 10.6%.

As Appendix C shows, more European students reached the upper stages of the Framework than Māori or Pasifika students and Māori students did better than Pasifika students. However, Asian students outperformed all other ethnic groups. This pattern is consistent with those found in other large-scale studies of mathematics achievement in which ethnicity has been a variable of interest, for example, TIMSS (see Garden, 1996, 1997), the Programme for International Assessment (PISA), and the National Education Monitoring Project (NEMP).

Appendix C also shows that students at low-decile schools did less well than those at medium- and high-decile schools. Although students at high-decile schools did better than those at medium-decile schools, the differences between students in the high- and medium-decile bands tended to be smaller than those between students at medium- and low-decile schools.

Appendix D shows the percentages of boys and girls at each Framework stage as a function of ethnicity. Each of the main ethnic groups showed a similar advantage for boys over girls at the upper stages of the Framework, and this pattern was consistent across all three operational domains:

addition and subtraction, multiplication and division, and proportion and ratio. Figure 1 shows the patterns for boys and girls from each main ethnic group on the domain of multiplication and division.

It is clear from Figure 1 that more boys than girls reach stage 7, advanced multiplicative part–whole. Figure 1 also shows that Asian and European students did better than Māori and Pasifika students. This can be seen in the greater percentages of students at stage 7 and the smaller percentages at stages 2–3 and 4, as well as by the fact that more of them were assessed on multiplication and division (shown by the height of the shaded bars; the blank space above the shaded bar represents the proportion who were *not* assessed on multiplication and division). It is interesting to note that more Māori boys than Māori girls reached the upper stages of the Framework. This pattern is opposite to that found for large-scale international comparisons such as TIMSS, where paper-and-pencil tests were used to assess mathematics achievement (see Garden, 1996, 1997). The TIMSS results showed that Māori girls tended to outperform Māori boys. It seems likely that the nature of the assessment is a crucial factor in determining these patterns. The diagnostic interview used in the NDP to assess students' mathematical proficiency (NumPA) involves the assessment of students individually by their own teachers, with tasks presented orally. Moreover, the emphasis is on the nature of the strategies used rather than simply whether or not the answer given was correct. By presenting tasks orally and expecting students to respond orally and to explain their thinking and reasoning, NumPA effectively minimises the literacy requirements and allows students to access the mathematics and demonstrate their mathematical proficiency unimpeded by literacy barriers. Of course, it is also possible that teachers unwittingly help certain students in the individual interview situation; that might help to explain the different patterns found for TIMSS and the NDP. However, evidence from Thomas, Tagg, & Ward (see p. 91 of this compendium) on the high level of agreement between the judgments of classroom teachers and those of independent researchers support the reliability of the individual interview data gathered with the NDP.

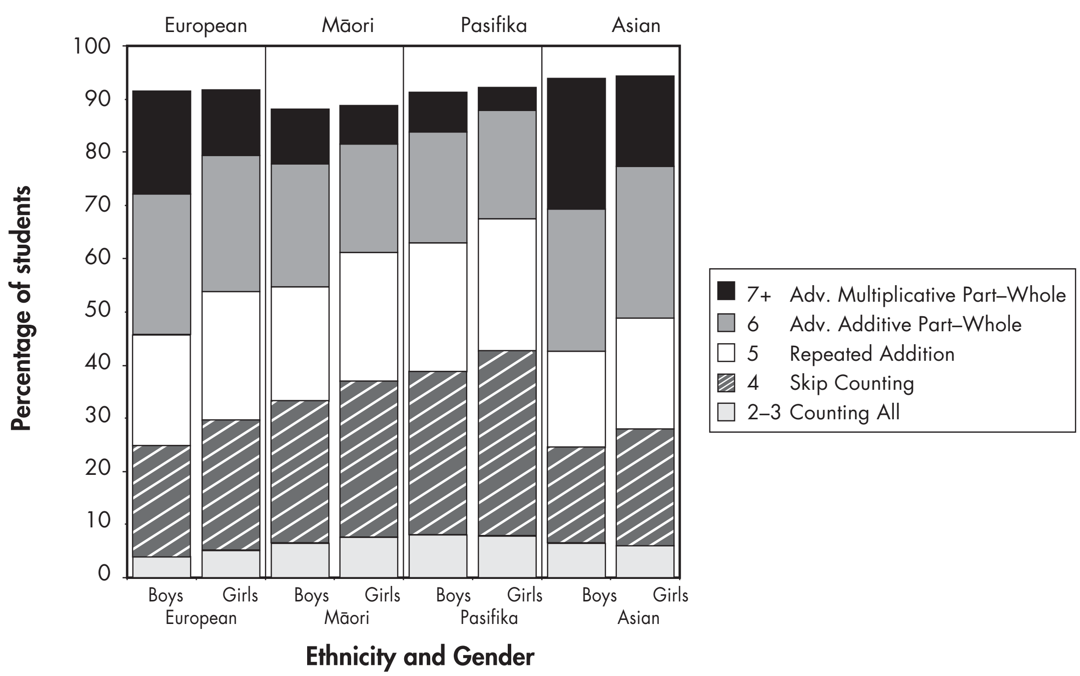

Figure 1. Percentages of students at each Framework stage on multiplication and division at the end of the year as a function of ethnicity and gender

Differential performance as a function of ethnicity: 2002–2005

It was interesting to look back at the patterns over time from 2002 to 2005 (see Table 1). The average Framework stage was calculated for each of the main ethnic groups – European, Māori, and Pasifika. At the beginning of 2003, the average Framework stage dipped. This was possibly because the criteria for crediting students with stage 6, advanced additive part–whole, was made stricter in 2003. Previously, students had been credited as being at stage 6 if they were "able to use a broad *range* of mental strategies to solve addition or subtraction problems with whole numbers". Although two different problems were given in 2002 (394 + 79 = □ and 403 – 97 = □), the instructions did not specify that students had to solve *both* problems correctly (see Ministry of Education, 2002). The instructions in 2003 stated that the students needed to be "able to use at least *two different* mental strategies to solve addition or subtraction problems with multi-digit numbers" (see Ministry of Education, 2003). As these were the only problems presented at that level, students needed to get both of them correct.

Table 1
Average Framework Stage on Addition and Subtraction for Year 0–8 Students 2002–2005

Year	Project Status	European	Māori	Pasifika	Asian
2002	Initial	4.05	3.78	3.50	4.30
	Final	4.61	4.33	4.06	4.79
2003	Initial	3.81	3.54	3.26	3.99
	Final	4.46	4.17	3.84	4.61
2004	Initial	3.93	3.77	3.61	4.06
	Final	4.57	4.38	4.23	4.76
2005	Initial	4.15	3.82	3.81	4.29
	Final	4.74	4.45	4.40	4.89

At the beginning of 2004, the average Framework stage was higher than it had been in 2003. This coincided with an increase in the proportion of students from high-decile schools (25.1% to 32.3%) and a corresponding decrease in the proportion of students from low-decile schools (37.6% to 28.9%). It is interesting to note that the differences between the three ethnic groups were smaller in 2004 than those during 2002 and 2003 (see Figure 1). At the beginning of 2005, the average Framework stage was again higher than at the start of the previous year. Once again, the proportion of students from high-decile schools had increased (32.3% to 37.2%), while those from low-decile schools had decreased (28.9% to 18.1%). It seems likely that the changes in the composition of the cohort might explain the increase in the average Framework stage in 2005.

Effect sizes

It is clear from Figure 1 that there were differences in the average Framework stage for students from different ethnic groups. The effect sizes for these differences were calculated to enable comparisons of effect sizes to be made. Effect sizes were calculated by dividing the average difference between two groups by the standard deviation for the two groups combined. Table 2 presents the average effect sizes for the comparisons between European and Māori and between European and Pasifika over the period 2002–2005.

According to Cohen's classification (see Fan, 2001), an effect size of 0.2 is considered "small" (a difference of less than a quarter of a standard deviation), those of 0.5 are thought to be "medium" (a difference of half a standard deviation), and those of 0.8 are considered "large" (a difference of more than three-quarters of a standard deviation). Hence, the effect sizes for the ethnicity comparisons are quite modest, particularly those for the European–Māori comparison (see later section in the paper on putting effect sizes into perspective).

Table 2
Average Effect Sizes for Ethnicity Comparisons (European vs Māori and Pasifika) 2002–2005

Year	European–Māori		European–Pasifika	
	Initial	Final	Initial	Final
2002	0.19	0.22	0.37	0.43
2003	0.17	0.22	0.35	0.46
2004	0.11	0.15	0.21	0.27
2005	0.22	0.22	0.23	0.26

Figure 2 shows the effect sizes for ethnicity differences over time. The pattern of effect sizes in Figure 2 shows a gradual reduction in the differences over time, apart from a slight rise for the European–Pasifika comparison at the end of the year, followed by a reduction in the magnitude of the differences in 2004 (both before and after the project). In 2005, the European–Pasifika differences appear to level out, while those for European–Māori comparisons increase. One possible explanation for the convergence of European-Māori and European-Pasifika differences in 2005 is that the proportion of year 7–8 students increased for European (20% to 23%) but decreased for Māori (28% to 24%), compared with the previous year. That is likely to have increased the difference in average Framework stage between European and Māori and hence to have increased the effect size. In 2004 (when the effect size for the European-Māori comparison was at its smallest), the proportion of year 7–8 Māori was at its highest.

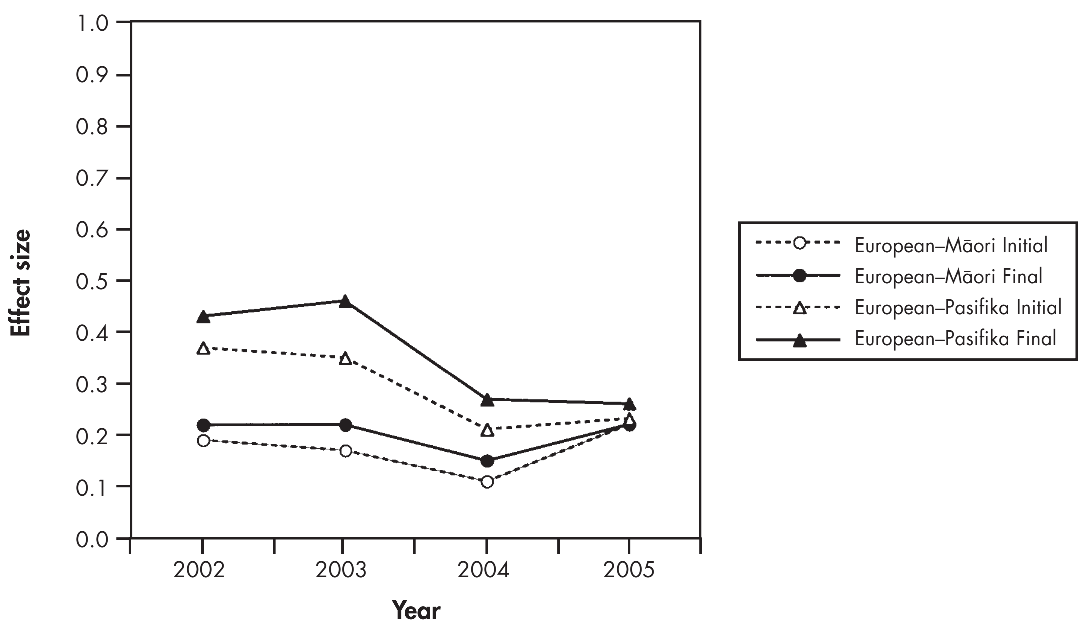

Figure 2. Effect sizes for differences in average Framework stage at the beginning (Initial) and end (Final) of the project over 2002–2005 as a function of ethnicity (European–Māori and European–Pasifika)

It is important to interpret cautiously the data that uses average Framework stage because of the problems already identified with the stages on the Framework not constituting an interval scale (because the steps at the lower end of the Framework are smaller than those at the upper end). A later section of this paper, which looks at patterns of progress with respect to identical starting points, provides a more reliable measure of students' performance and progress.

Patterns of Progress

Patterns of progress were examined by looking at the proportions of students who moved to a higher Framework stage relative to particular starting points. Appendix E shows the percentages of students at each initial stage who moved to a higher Framework stage. Separate results are shown for European, Māori, and Pasifika, and results are presented for 2002 to 2005. Figures 3 to 6 show the graphs presenting the patterns of progress for European, Māori, and Pasifika students over 2002 to 2005. Students who started at stage 0, emergent, or stage 1, one-to-one counting, showed the greatest progress, with about 80% of students moving to a higher Framework stage.

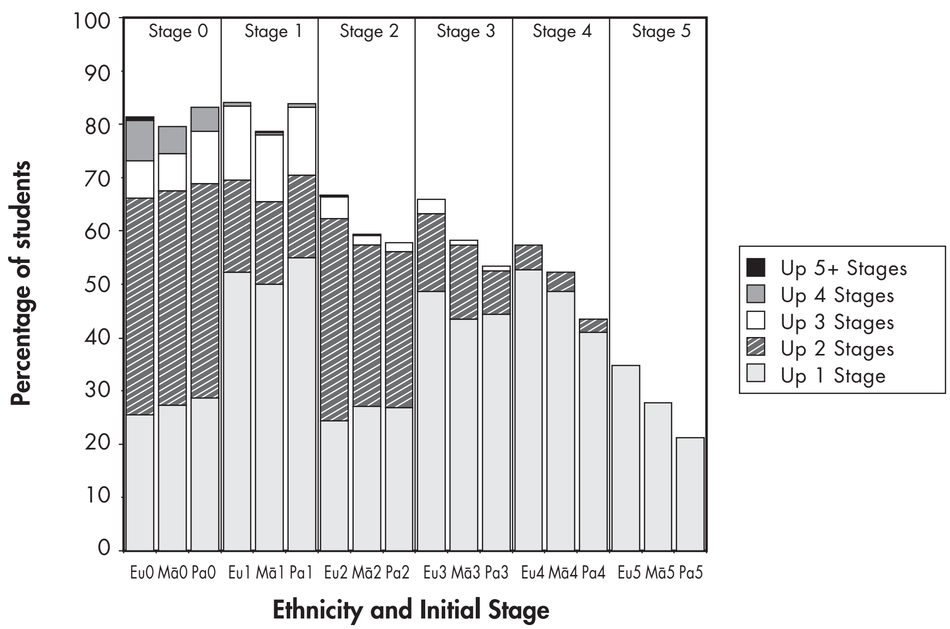

Figure 3. Percentage of students who progressed to a higher Framework stage on addition and subtraction as a function of ethnicity and initial stage (2002)

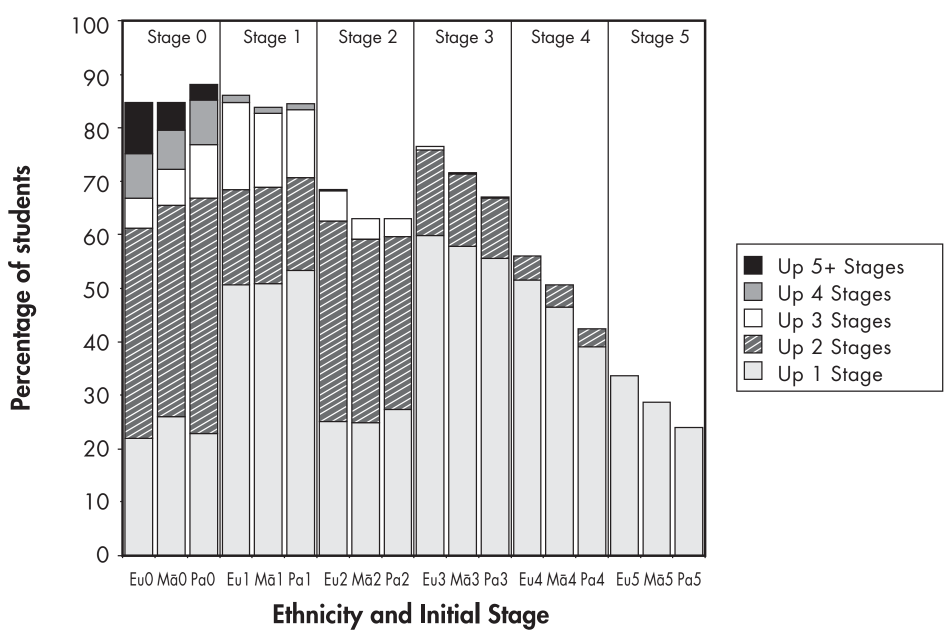

Figure 4. Percentage of students who progressed to a higher Framework stage on addition and subtraction as a function of ethnicity and initial stage (2003)

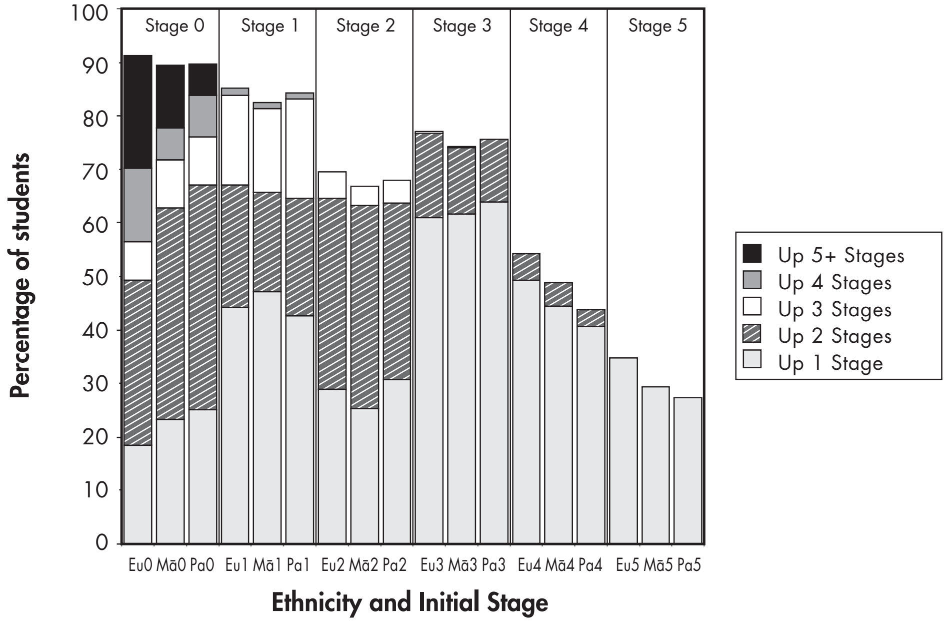

Figure 5. Percentage of students who progressed to a higher Framework stage on addition and subtraction as a function of ethnicity and initial stage (2004)

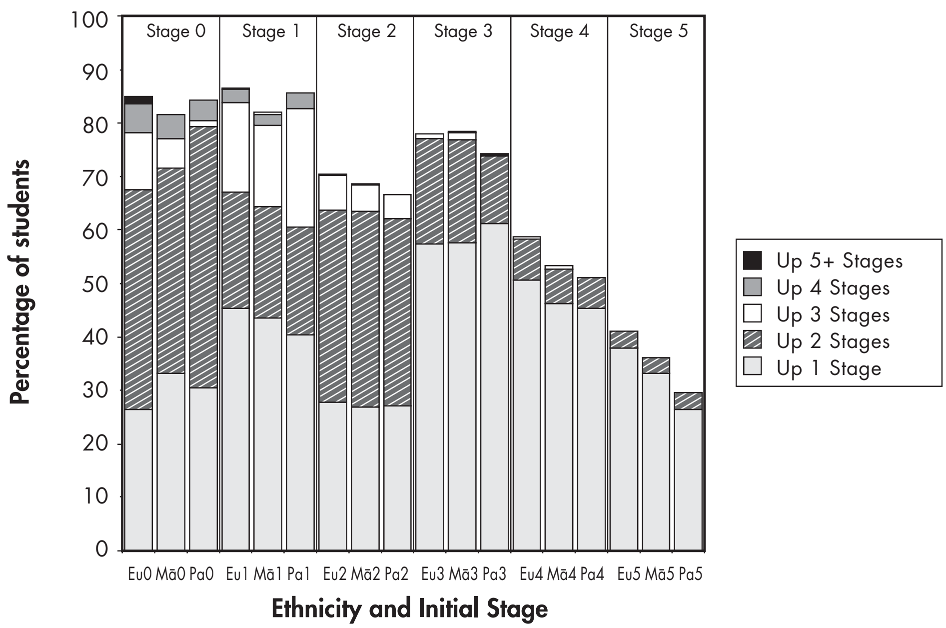

Figure 6. Percentage of students who progressed to a higher Framework stage on addition and subtraction as a function of ethnicity and initial stage (2005)

In some cases, Pasifika students made greater progress than European or Māori students. Approximately two-third of students who began at stage 2, counting from one with materials, or stage 3, counting from one with imaging, moved to a higher Framework stage. Progress was better for those who started at stage 3 than for those who started at stage 2, despite the fact that stage 3 students could progress only three stages at the most, whereas those at stage 2 could potentially improve four stages. This suggests that, once students understand how to use counting to work out the total when two collections are joined, they make rapid progress through at least stages 2 and 3. In the earlier years of the project (2002 & 2003), European students made substantially better progress than Māori and Pasifika students. However, in more recent years (2004 and 2005), the differences in progress between the three ethnic groups have substantially reduced.

At stage 4, advanced counting, and stage 5, early additive part–whole, European students tended to make greater progress than Māori students, who in turn made more progress than Pasifika students. However, there is a steady pattern of improvement over time for both Māori and Pasifika students. For example, for Māori students who began the project at stage 5, the percentages moving to a higher Framework stage (that is, stage 6, advanced additive part–whole) have increased from 27.8% in 2002 to 36.2% in 2005. The corresponding figures for Pasifika students were 21.3% in 2002 to 29.7% in 2005. Patterns of progress for European students have tended to remain fairly stable (34.7%, 33.7%, 34.8% for 2002, 2003, and 2004 respectively), apart from 2005, when there was a sudden increase to 41.1%. Similarly, for those students who started at stage 4, the proportion going up two stages to stage 6 has increased steadily over the years between 2002 and 2005. For example, the percentages of Māori students at stage 4 who went up to stage 6 has improved from 3.7% in 2002 to 4.2%, 4.6%, and 7.0% in 2003, 2004, and 2005 respectively. The corresponding percentages for Pasifika were 2.5%, 3.2%, 3.3%, and 5.9%. The percentages of Pasifika students at stage 3 who went up to a part–whole stage (either stage 5, early additive part–whole, or stage 6, advanced additive part–whole) increased from 9.0% in 2002 to 11.4%, 11.8%, and 13.0% in 2003, 2004, and 2005 respectively. Although these changes are quite small, they could be interpreted as indicating that the project is becoming increasingly effective in enhancing the mathematical proficiency of Māori and Pasifika students. Alternatively, improvements for these two groups could be explained by the greater proportions of Māori and Pasifika students in high-decile schools in more recent years.

Appendix F shows the final Framework stage achieved as a function of initial stage for European, Māori, and Pasifika students and for students in high-decile and low-decile schools. This table allows comparison for a particular group across time (horizontal) and comparison between groups who started at particular stages in a particular year (vertical).

It is important to note here that it is difficult to examine patterns of progress on the multiplication and division and proportion and ratio domains because so many students were not given tasks from these domains, particularly at the time of the initial assessment. Hence, with no initial data, there is no baseline from which to examine progress. This applies to all students assessed on Form A of NumPA, which does not include multiplication and division or proportion and ratio tasks.

Narrowing the achievement gap

In order to investigate the extent to which the NDP narrowed the gap in mathematics achievement between European students and Māori and Pasifika groups, effect sizes were calculated for the differences between European and Māori. European and Pasifika, and students in high-decile and low-decile schools for 2002 to 2005. Because of the problems with the Framework stages not constituting an interval scale, separate effect sizes were calculated for students who began the projects at each initial stage. The median effect size was then used as an indicator of the pattern overall (see Appendix G). Table 3 presents the median effect sizes for each comparison between 2002 and 2005. Analysis shows that the median effect size for differences between European and Māori students reduced from 0.14 in 2002 to 0.09 in 2005. The corresponding values for Pasifika students reduced from 0.26 in 2002 to 0.17 in 2005. Differences between students in the high-decile and low-decile bands reduced from 0.21 in 2002 to 0.13 in 2005.

The effect-size analysis was limited to the addition and subtraction domain because that is the only strategy domain on which there is complete data. The domains of multiplication and division and proportion and ratio are problematic for two reasons: not all students were given the opportunity to do tasks from these other two domains and some students were only given a chance to do tasks from these two domains at the end of the year but not the beginning, so there is no data about their initial stage. Hence, this data is not included in this paper.

Table 3
Median Effect Sizes for Comparisons of Progress on Addition and Subtraction for Students Who Started at Identical Framework Stages (European vs Māori and Pasifika, high-decile vs low-decile) 2002–2005

Year	European–Māori	European–Pasifika	High–Low decile
2002	0.14	0.26	0.21
2003	0.13	0.21	0.22
2004	0.10	0.16	0.14
2005	0.09	0.17	0.13

The changes in median effect size for the three comparisons on addition and subtraction are shown in Figure 7.

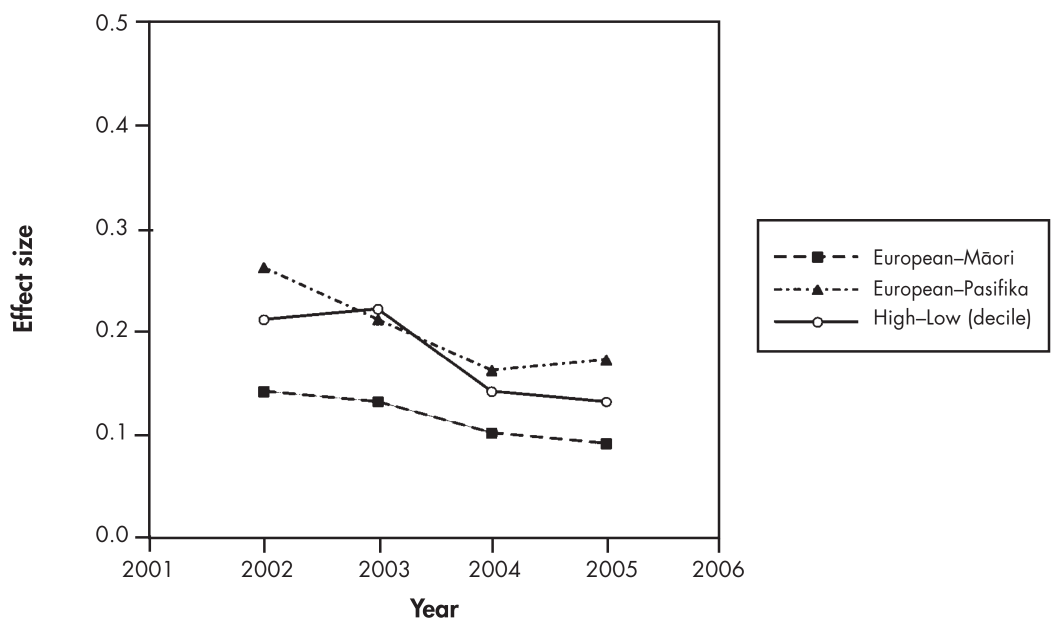

Figure 7. Changes in median effect size for differences in progress on addition and subtraction as a function of ethnicity (European–Māori and European–Pasifika) and decile (high–low)

Figure 7 shows that the median effect size for differences between European and Māori, European and Pasifika, and students from high-decile and low-decile schools reduced over the years between 2002 and 2005, although there is some levelling out for the European–Pasifika comparison. It is difficult to know whether this reduction in effect size is the result of changes in the composition of the cohort, with a greater percentage of students coming from higher decile schools in later years, or whether this reflects the improved effectiveness of the NDP initiative itself.

Interrelationships between multiplicative thinking and other domains

Despite the best efforts of numeracy facilitators, the percentage of students who reached stage 7, advanced multiplicative part–whole, was disappointing (a third of year 8 students). An analysis of the interrelationships between various domains was completed using final interview data (in order to maximise the reliability of the assessment data), and the data was aggregated across the years 2002 to 2005. Across those four years of the project, more than 36 000 year 0–8 students reached stage 7 by the end of the year. The analysis was completed for each of these stage 7 groups (referred to as stage 7 M/D to distinguish them from students who reached stage 7 on other domains), to see how these students performed on other key domains of the Number Framework compared with those who hadn't yet reached stage 7 (see Appendix H). The data was then aggregated by averaging the percentages across the four years 2002–2005 (see Appendix H and Appendix I). Marked differences between stage 7 M/D students and those below stage 7 on the multiplication and division domain were found on proportion and ratio, fractions, and basic facts. As Appendix H shows, 71% of stage 7 M/D students had reached at least stage 7 on proportion and ratio, compared with only 14% of stage 6 M/D and 2% of stage 5 M/D students. Stage 7 on proportion and ratio can be achieved by working out that, if 12 is $\frac{2}{3}$ of a number, the number must be 18. In order to do this task successfully, students need to understand how to divide the 12 in half to work out how much one-third is, then multiply this number by 3 to work out what three-thirds are altogether.

A similar difference was found on fractions, where 53% of stage 7 M/D students were able to recognise equivalent fractions ($\frac{2}{3}$ and $\frac{6}{9}$) while ordering fractions with unlike numerators and denominators. Only 9% of stage 6 M/D students and only 1% of stage 5 M/D students could do this. On basic facts, almost 68% of stage 7 M/D students could recall division facts and some could also recall common factors and multiples, compared to only 20% of stage 6 M/D students and 4% of stage 5 M/D students.

Appendix I shows the percentage of students at each of the upper levels on the proportion and ratio domain who were at particular levels on other domains. Only 65% of students who were at stage 7, advanced multiplicative, on proportion and ratio were at stage 7 on multiplication and division, suggesting that the fraction problem ("If 12 is $\frac{2}{3}$ of a number, what is the number?") was an easier task than the multiplication and division problems used to assess stage 7 on the multiplication and division domain. This is borne out by the fact the percentages of students at stage 7 P/R were lower on fractions, place value, and basic facts than those found for students at stage 7 M/D.

The results suggest a close connection between advanced multiplicative thinking and fractions and division. It may be that teachers have focused on multiplication without also building a strong understanding of division. This could explain why the percentage of students reaching stage 7, advanced multiplicative, is only about 34%. Working with fractions requires a reasonable understanding of division also, and this may be the reason that fraction knowledge and fraction strategies (as seen at stage 7 on the proportion and ratio domain) seem to be related to multiplication and division. Further work with division may help students with their understanding of fractions. It is also possible that further work with fractions, particularly encouraging students to *understand* fractions rather than engage in mindless manipulation of digits, might also foster the development of multiplicative reasoning.

Introduction of new stage 7 A/S and stage 8 M/D tasks in 2005

Appendix J shows the performance of students who are at stage 7 at the end of the year on other domains of the Number Framework that were assessed at the same time. It is interesting to compare the students at stage 7 A/S with those at stage 7+ M/D and stage 7+ P/R. The largest number of students (6612) was at stage 7 or higher on multiplication and division (with whole numbers), and the smallest number (2742) was at stage 7 on addition and subtraction (addition and subtraction with decimals and fractions). The reason for the smaller number at stage 7 on addition and subtraction may have been that many students were not familiar with decimals. There were almost as many students at stage 7 or higher on proportion and ratio (6250) as at stage 7 or higher on multiplication and division (6612).

Of the students who were assessed as being at stage 7, whether it was on addition and subtraction, multiplication and division, or proportion and ratio, virtually all were at stage 5 or above on all of the other strategy domains. More than 90% of them were at stage 6 or higher on all of the other strategy domains. It is interesting to consider what might have kept those students at stage 6 rather than being rated at stage 7. On the proportion and ratio tasks, the method used to solve $\frac{3}{4}$ of 28 was crucial in determining whether the student was at stage 6 or stage 7. It was the use of addition and multiplication (for example, finding $\frac{1}{4}$ of 28 by going up in 4s from 20, as in "$\frac{1}{4}$ of 20 is 5, so $\frac{1}{4}$ of 24 is 6, so $\frac{1}{4}$ of 28 is 7", then multiplying by 3) rather than division and multiplication ("divide 28 by 4 then multiply that number by 3") that led to a stage 6 judgment. Included in this group were students who used successive halving to find the quarter ("$\frac{1}{2}$ of 28 is 14, and $\frac{1}{2}$ of 14 is 7"), and then multiplied by 3. To be at stage 7, students also needed to solve the problem that involved using the information that $\frac{2}{3}$ of a number is 12 to work out what the number is. On the multiplication and division tasks, 25.3% of the stage 7 P/R students were rated at stage 6 rather than stage 7. This would have been because they were not able to use at least two different advanced mental strategies to solve the 24 x 6 task and the 72 ÷ 4 task. It would be very interesting to know which of the two tasks the students had the most problems with. There are part–whole strategies that can be used to solve the multiplication task that are very similar to the written algorithm for multiplication (for example, first multiply 6 by 4, then multiply 6 by 20 and add the two products). The division task is not as closely connected to the "long division" algorithm (for example, build up to 72 in parts such as 4 x 10 and 4 x 8, use successive halving as in "$\frac{1}{2}$ of 72 is 36 and $\frac{1}{2}$ of 36 is 18", or a compensation strategy like "4 x 20 is 80; 80 minus 2 x 4 is 72").

Almost 88% of stage 7 A/S students were at stage 7 or higher on multiplication and division (87.6% could do multiplication and division with whole numbers, and 40.2% could do so with decimals), whereas 80% of them were at stage 7 or higher on proportion and ratio (80.2% worked out what a number was if $\frac{2}{3}$ of it was 12, and 27.9% could solve the two ratio problems – wool and mittens and the percentage of boys in Ana's class). It was interesting to note that 68% of the 6612 students at stage 7 or higher on multiplication and division were at stage 7 or higher on proportion and ratio (n = 4489). This corresponded almost exactly to the 72% of the 6250 students at stage 7 or higher on proportion and ratio who were at stage 7 or higher on multiplication and division (n = 4488). These figures indicate how complicated the relationships are between the various tasks, levels, and domains.

More of the stage 7 A/S students were at stage 7 on fractions (73.9%) compared with the stage 7+ M/D students (53.1%) and stage 7+ P/R students (52.0%). This reflects the fact that the stage 7 A/S tasks were the hardest and the stage 7+ M/D tasks were the easiest. Stage 7 A/S students were also better at basic facts (80.7% were at stage 7) compared to the stage 7+ M/D students

(66.4%) and the stage 7+ P/R students (66.5%). Hence, it was possible to order the tasks used to assess stage 7 by difficulty level. The easiest task seemed to be the multiplication and division tasks with whole numbers ($24 \times 6 = \square$ and $72 \div 4 = \square$). However, it may be that the size of the numbers also had an impact on the difficulty level of the task, but there were not enough different tasks to allow comparisons to be made so that that issue could be explored. The fractions task used to assess stage 7 on proportion and ratio ("If 12 is $\frac{2}{3}$ of a number, what is the number?") was of medium difficulty. The hardest tasks by far appeared to be the addition and subtraction tasks with fractional number ($2 - (\frac{3}{4} + \frac{7}{8}) = \square$ and $5.3 - 2.89 = \square$). It is difficult to know whether it was the fractions task or the decimals task that was the harder of the two, but it seems likely that decimals presented the greatest difficulty.

Appendix K shows the performance of students on stage 8 M/D and stage 8 P/R on other domains on the Number Framework assessed at the same time. It was possible to compare the performance of students at stage 8 M/D with those at stage 8 P/R. The two ratio tasks (wool and mittens, $10{:}15 = \square{:}6$, and the percentage of boys in Ana's class, $21{:}35 = \square{:}100$) were slightly harder (n = 1008) than division with decimals and whole numbers ($2.4 \div 0.15 = \square$ and $26 \div 8 = \square$) (n = 1444). Similar proportions of students at stage 8 M/D were at stage 7 on addition and subtraction compared with those at stage 8 P/R (76.1% vs 75.7%). Their performance on fractions was also similar (85.1% of stage 8 M/D and 90.0% of stage 8 P/R were at stage 7+). Likewise, similar proportions were at stage 7+ on basic facts (88.6% and 89.9% of stage 8 M/D and stage 8 P/R respectively).

It was interesting to look at how the students rated at stage 7 on the various operational strategy domains did on fractional number – one of the knowledge domains. Virtually all of the stage 7 students were able to order unit fractions, so it seems they did understand the idea that the larger the denominator, the smaller the fraction. A notable proportion of those at stage 7+ M/D (20.5%) and stage 7+ P/R (18.0%) were unable to coordinate numerators and denominators (as in recognising that $\frac{8}{6}$ is the same as $1\frac{2}{6}$ or $1\frac{1}{3}$), whereas only 8.5% of stage 7 A/S students were in that category. A quarter of stage 7+ M/D and P/R students were able to coordinate numerators and denominators but did not recognise the equivalence of $\frac{2}{3}$ and $\frac{6}{9}$ when ordering fractions with different denominators. Only 16.1% of stage 7 A/S students were in that category.

The differences in performance on identical stages on different domains raise some important issues about what it means to be a particular Framework stage on a particular domain. It seems likely also that the difficulty level of the tasks is affected by the size of the numbers as well as the kind of operation and that caution is needed about drawing firm conclusions about Framework stages on the basis of a student's performance on just one or two tasks.

The difficulties with fractions have been well documented in the literature (for example, Behr et al., 1983; Charalambous & Pitta-Pantazi, 2005; Verschaffel, Greer, & Torbeyns, 2006). According to these writers, fractions are very difficult to teach and to learn because the concept of fractions consists of several sub-constructs and understanding fractions requires an understanding of each of the sub-constructs as well as the ways in which the sub-constructs are connected. Underpinning each of the four sub-constructs of fraction understanding is the idea of part–whole comparison or partitioning. Unlike partitioning for addition and subtraction, which can be of unequal parts, partitioning for multiplication, division, and fractions must be of equal-sized parts. According to Behr et al. (p. 93), many student difficulties in algebra can be traced back to an incomplete understanding of earlier fraction ideas.

The sub-constructs include *ratio* (the idea of relative magnitude, necessary for understanding ideas about proportion and equivalence, as in $\frac{3}{4}$ is the same as $\frac{6}{8}$), *operator* (necessary for multiplication of fractions, as in $\frac{3}{4}$ of 10 metres), *quotient* (necessary for problem-solving, as in $\frac{1}{4}$ of 20 means the division of 20 by 4), and *measure* (necessary for addition of fractions, as in $\frac{3}{4}$ is the same as $\frac{1}{4} + \frac{1}{4} + \frac{1}{4}$). Another way of looking at fractions is to see them as both a process (for example, $\frac{3}{4}$ involves division of 3 by 4) and a product ($\frac{3}{4}$ is the result of dividing 3 by 4) (see Verschaffel et al., 2006). Students can easily confuse fractions as *numbers* (as in $\frac{3}{4}$ of a metre) and fractions as *operations* (as in $\frac{3}{4}$ of 10 metres). Fractions appear in two places in the Number Framework: as part of the operational strategy domain of proportion and ratio as well as in the knowledge domain of fractional number. A lot more work is needed to investigate how the various aspects of fractional number fit within the Number Framework. More work also needs to be done on ways of helping students understand fractional number better.

Putting effect sizes into perspective

The NDP was initially designed to raise mathematics achievement for all students. The various projects seem to have been fairly successful at doing this, despite some concerns about how far on the Framework students are progressing on domains such as multiplication and division and proportion and ratio. However, with such a large amount of data, the opportunity to look at patterns of performance and progress as a function of variables such as ethnicity and school decile is too good to let pass. These analyses have shown that, although all students made progress, the gaps in the achievement of European and Asian students compared with Māori and Pasifika students continued to exist. In general, the achievement gaps between groups seem to have reduced slightly over the years, but this may have been due to changes in the composition of the sample. It is important to see these differences in the wider perspective. When the effect sizes for these differences are compared with corresponding differences found on other large-scale studies of mathematics achievement, it becomes clear that the effect sizes for the differences on the NDP are substantially smaller than those found in the other studies. For example, on the TIMSS study, effect sizes are about three-quarters of a standard deviation for the European–Māori comparison and about one standard deviation for the European–Pasifika comparison. Based on Cohen's classification (see Fan, 2001), these are "large" effect sizes (that is, about 0.8 or more), whereas those on the NDP are mostly about 0.2, which is considered "small" on Cohen's classification. The effect sizes for the PISA study are smaller than those on TIMSS (0.38 and 0.53), but this study differs in an important way from the others. The PISA study looked at students aged between 15 years 3 months and 16 years 2 months. There is ample evidence from educational statistics that many Māori and some Pasifika students have left school by the age of 15 years, or even earlier. Hence, the comparison does not include a full cohort of students. It is often those students who are not succeeding at secondary school who decide to leave early. Hence, the PISA results do not include the full range of mathematics achievement levels, and this will inevitably have somewhat reduced the magnitude of effect sizes. Table 4 shows the effect sizes for each study/group, and indicates that the disparities in achievement between ethnic groups are not as great as was previously thought.

Table 4
Effect Sizes for the Comparison of European vs Māori and European vs Pasifika on TIMSS, PISA, and the NDP

		Comparison	
		European–Māori	European–Pasifika
TIMSS	Yr 5 1994	0.73	0.95
TIMSS	Yr 5 1998	0.65	0.97
TIMSS	Yr 9 1994	0.71	1.15
TIMSS	Yr 9 1998	0.66	0.96
PISA	15 yrs 2000	0.38	0.53
NDP	Initial 2002	0.19	0.37
NDP	Initial 2003	0.17	0.35
NDP	Initial 2004	0.11	0.21
NDP	Initial 2005	0.22	0.23

Conclusions

As the amount of data collected in the course of the NDP grows, many more questions are raised about the nature of the Number Framework and the interrelationships among the domains. Not only does the Number Framework provide teachers with a valuable structure to help organise their assessment and teaching, it is also a valid area of enquiry in its own right. It is hoped that, as the emphasis of the NDP shifts away from the primary level towards the intermediate and secondary levels, this data-gathering process continues to provide researchers with valuable information that can help throw further light on the complexities of the Framework itself.

The data indicate that there are important issues to investigate further with respect to multiplicative thinking and understanding of fractional numbers. There are now three different sets of tasks to explore multiplicative thinking: addition and subtraction, multiplication and division, and proportion and ratio tasks. Students' performance varied according to which of these was being assessed. There is clearly an important link between understanding division and fractional number, and this link needs to be explored further.

References

Behr, M. J., Lesh, R., Post, T. R., & Silver, E. A. (1983). Rational-number concepts. In R. Lesh & M. Landau (Eds), *Acquisition of mathematics concepts and processes.* (pp. 91–126). London: Academic Press.

Bobis, J., Clarke, B., Clarke, D., Thomas, G., Wright, R., Young-Loveridge, J., & Gould, P. (2005). Supporting teachers in the development of young children'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: British Columbia Ministry of Education.

Charalambous, C. Y. & Pitta-Pantazi, D. (2005). Revisiting a theoretical model on fractions: Implications for teaching and research. In H. L. Chick & J. L. Vincent (Eds), *Proceedings of the 29th PME International Conference, 2,* 233–240.

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

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

Garden, R. A. (1996). *Mathematics performance of New Zealand form 2 and form 3 students: National results from New Zealand's participation in the Third International Mathematics and Science Study.* Wellington: Ministry of Education.

Garden, R. A. (1997). *Mathematics and science performance in middle primary school: Results from New Zealand's participation in the Third International Mathematics and Science Study.* Wellington: Ministry of Education.

Ministry of Education (2001). *Curriculum Update, 45: The numeracy story.* Wellington: Learning Media.

Ministry of Education (2003). *Book 2: The diagnostic interview.* Wellington: Ministry of Education.

Ministry of Education (2004). *Book 2: The diagnostic interview.* Wellington: Ministry of Education.

Ministry of Education (2002). *The diagnostic interview.* Wellington: Ministry of Education.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: NCTM.

New South Wales Department of Education and Training (2001). *Count Me In Too: Professional development package.* Sydney: NSW Department of Education and Training.

Thomas, G., Tagg, A., & Ward, J. (2006). Numeracy assessment: How reliable are teachers' judgments? In this volume.

Verschaffel, L., Greer, B., & Torbeyns, J. (2006). Numerical thinking. In A. Gutierrez & P. Boero (Eds), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 51–82). Rotterdam: Sense Publishers.

Young-Loveridge, J. (2005). Patterns of performance and progress: Analysis of 2004 data. In *Findings from the New Zealand Numeracy Development Project 2004* (pp. 5–20 and pp. 115–127). Wellington: Ministry of Education.