

# 8 Performance of Subgroups

Although national monitoring has been designed primarily to present an overall national picture of student achievement, there is some provision for reporting on performance differences among subgroups of the sample. Eight demographic variables are available for creating subgroups, with students divided into subgroups on each variable, as detailed in Chapter 1 (p5).

Analyses of the relative performance of subgroups used the total score for each task, created as described in Chapter 1 (p5).

## SCHOOL VARIABLES

Five of the demographic variables related to the schools the students attended. For these five variables, statistical significance testing was used to explore differences in task performance among the subgroups. Where only two subgroups were compared, differences in task performance between the two subgroups were checked for statistical significance using t-tests. Where three subgroups were compared, one-way analysis of variance was used to check for statistically significant differences among the three subgroups.

Because the number of students included in each analysis was quite large (approximately 450), the statistical tests were quite sensitive to small differences. To reduce the likelihood of attention being drawn to unimportant differences, the critical level for statistical significance for tasks reporting results for individual students was set at  $p = .01$  (so that differences this large or larger among the subgroups would not be expected by chance in more than one percent of cases). For tasks administered to teams or groups of students,  $p = .05$  was used as the critical level, to compensate for the smaller numbers of cases in the subgroups.

For the first four of the five school variables, statistically significant differences among the subgroups were found for slightly less than 16 percent of the tasks at both year levels. For the remaining variable, statistically significant differences were found on nearly two thirds of the tasks at both levels. In the detailed report below, all

differences mentioned are statistically significant (to save space, the words “statistically significant” are omitted).

### School Size

Results were compared from students in larger, medium size, and small schools (exact definitions were given in Chapter 1 (p5).

For year 4 students, there were differences among the three subgroups on two of the 64 tasks. Students attending small schools scored lowest on *Number Facts (Multiplication)* (p13) and on *Link Task 5* (p29). There were no differences on questions of the *Mathematics Survey* (p55).

For year 8 students there were differences among the three subgroups on one of the 91 tasks. Students from medium size schools scored highest on *Link Task 42* (p49). There were no differences on questions of the *Mathematics Survey* (p55).

### Community Size

Results were compared for students living in communities containing over 100,000 people (main centres), communities containing 10,000 to 100,000 people (provincial cities) and communities containing less than 10,000 people (rural areas).

For year 4 students, there were differences among the three subgroups on six of the 64 tasks. Students from provincial cities scored lowest and students from main centres scored highest on five of these tasks: *Algorithms (Division)* (p14), *Number Facts (Multiplication)* (p13), *Link Task 3* (p29), *Link Task 12* (p29)



and *Link Task 13* (p30). Students from main centres scored highest and students from rural areas scored lowest on the remaining task, *Algorithms (Subtraction)* (p14). There were no differences on questions of the *Mathematics Survey* (p55).

For year 8 students, there was a difference among the three subgroups on one of the 91 tasks. Students from provincial cities scored lowest on *Link Task 22* (p30). There were no differences on questions of the *Mathematics Survey* (p55).

### School Type

Results were compared for year 8 students attending full primary and intermediate (or middle) schools, and students attending year 7 to 13 high schools.

In comparing students attending full primary and intermediate (or middle) schools, there were statistically significant differences on three of the 91 tasks. Students attending full primary schools scored higher than students attending intermediate (or middle) schools on *Thermometer* (p38) and *Link Task 19* (p30). Students attending intermediate (or middle) schools scored higher than students attending full primary schools on *Link Task 20* (p30). There was one difference on the questions of the *Mathematics Survey* (p55). Students attending full primary schools reported significantly higher ratings for the item, “How much do you like doing maths in your own time?” as compared to the students attending intermediate (or middle) schools.



In comparing students attending intermediate (or middle) schools to those attending year 7 to 13 high schools, there were statistically significant differences on six of the 91 tasks. Students attending year 7 to 13 high schools scored higher than students attending intermediate (or middle) schools on all six tasks: *Numbers on Lines* (p23), *Equivalents* (p28), *Thermometer* (p38), *Awesome Angles* (p48), *Link Task 6* (p29) and *Link Task 39* (p49). There were no differences on questions of the *Mathematics Survey* (p55).

### Zone

Results achieved by students from Auckland, the rest of the North Island, and the South Island were compared.

For year 4 students, there were differences among the three subgroups on nine of the 64 tasks. Students from the Auckland scored highest on 7 tasks: *Number Facts (Multiplication)* (p13), *Algorithms (Division)* (p14), *Page of Stamps* (p16), *Number Patterns* (p19), *Fractions* (p24), *Link Task 3* (p29) and *Link Task 11* (p29). Students from the South Island scored highest on the remaining two tasks: *Letter* (p34) and *How Much Change?* (p34). Students from the South Island scored lowest on two tasks: *Number Facts*

*(Multiplication)* (p13) and *Link Task 11* (p29); students from the rest of the North Island scored lowest on all remaining tasks. There was one difference on the questions of the *Mathematics Survey* (p55). Students from Auckland were most positive and students from the South Island were least positive on the question, “How do you feel about doing things in maths you haven’t tried before?”

For year 8 students, there were differences among the three subgroups on seven of the 91 tasks. Students from the South Island scored highest on six tasks: *Fractions* (p24), *Change* (p34), *Nets* (p46) *Pick A Teddy* (p51), *Link Task 43* (p49), and *Link Task 44* (p52). Students from the rest of the North Island scored highest on the remaining task, *Tangram* (p23). Students from Auckland scored lowest on six tasks: *Tangram* (p23), *Fractions* (p24), *Change* (p34), *Nets* (p46), *Pick a Teddy* (p51) and *Link Task 44* (p52). Students from the rest of the North Island scored lowest on the remaining task, *Link Task 43* (p49). There was one difference on the questions of the *Mathematics Survey* (p55). Students from the South Island were most positive and students from Auckland were least positive on the question, “How much do you like doing maths in your own time?”

### Socio-Economic Index

Schools are categorised by the Ministry of Education based on census data for the census mesh blocks where children attending the schools live. The resulting index takes into account household income levels and categories of employment. It uses 10 subdivisions, each containing 10 percent of schools (deciles 1 to 10).

For our purposes, the bottom three deciles (1-3) formed the low decile group, the middle four deciles (4-7) formed the medium decile group and the top three deciles (8-10) formed the high decile group. Results were compared for students attending schools in each of these three groups.

For year 4 students, there were differences among the three subgroups on 40 of the 64 tasks. Because of the number of tasks involved, the specific tasks are not listed here. In each case, performance was lowest for students in the low decile group. Students in the high decile group performed better than students in the medium decile group on all but five tasks; however, these differences were quite small. There were significant differences on three of the questions on the *Mathematics Survey* (p55). Students in the low decile group were more positive than students in the high decile group on two questions: “How much do you like doing maths on your own?” and “How much do you like doing maths with others?” Students in the low decile group were more positive than students in the high and middle decile groups on the question, “How much do you like doing maths in your own time?”

For year 8 students, there were differences among the three subgroups on 59 of the 91 tasks. Because of the number of tasks involved, the specific tasks are not listed here. In each case, performance was lowest for students in the low decile group. Students in the high decile group performed better than students in the medium decile group on all but two tasks; however, these differences were quite small. There were no differences among groups on the questions of the *Mathematics Survey* (p55).

## STUDENT VARIABLES

Three demographic variables related to the students themselves:

- *Gender*: boys and girls
- *Ethnicity*: Māori, Pasifika and Pakeha (this term was used for all other students)
- *Language used predominantly at home*: English and other.

During the cycle of the Project that took place from 1999-2002, special supplementary samples of students from schools with at least 15 percent Pasifika students enrolled were included. These allowed the results

of Pasifika students to be compared with those of Māori and Pakeha students attending these schools. By 2002, with Pasifika enrolments having increased nationally, it was decided that from 2003 onwards a better approach would be to compare the results of Pasifika students in the main NEMP samples with the corresponding results for Māori and Pakeha students. This gives a nationally representative picture, with the results more stable because the numbers of Māori and Pakeha students in the main samples are

much larger than their numbers previously in the special samples.

The analyses reported compare the performances of boys and girls, Pakeha and Māori students, Pakeha and Pasifika students, and students from predominantly English-speaking and non-English-speaking homes.

For each of these three comparisons, differences in task performance between the two subgroups are described using effect sizes and statistical significance.





For each task and each year level, the analyses began with a t-test comparing the performance of the two selected subgroups and checking for statistical significance of the differences. Then the mean score obtained by students in one subgroup was subtracted from the mean score obtained by students in the other subgroup, and the difference in means was divided by the pooled standard deviation of the scores obtained by the two groups of students. This computed effect size describes the magnitude of the difference between the two subgroups in a way that indicates the strength of the difference and is not affected by the sample size. An effect size of +.30, for instance, indicates that students in the first subgroup scored, on average, three tenths of a standard deviation higher than students in the second subgroup.

For each pair of subgroups at each year level, the effect sizes of all available tasks were averaged to produce a mean-effect size for the curriculum area and year level, giving an overall indication of the typical performance difference between the two subgroups.

### Gender

Results achieved by male and female students were compared using effect-size procedures.

For year 4 students, the mean-effect size across the 63 tasks was .08 (boys averaged 0.08 standard deviations higher than girls). This difference is small. There were statistically significant differences ( $p < .01$ ) favouring boys on eight of the 63 tasks: *Algorithms (Subtraction)* (p14), *12 Bears* (p17), *How Much Change?* (p34), *Link Task*

*5* (p29), *Link Task 9* (p29), *Link Task 10* (p29), *Link Task 11* (p29) and *Link Task 30* (p42). There were differences on two questions of the *Mathematics Survey* (p55). Boys were more positive than girls for the question, “How good does your teacher think you are at maths?” and girls were more positive than boys in response to the question, “How much do you like doing maths in your own time?”

For year 8 students, the mean-effect size across the 89 tasks was .03 (girls averaged 0.03 standard deviations higher than boys); this is a small difference. There were statistically significant differences on seven of the 89 tasks, with girls performing better on all seven tasks: *Letter* (p34), *Snacks* (p38), *Trapezium* (p45), *Link Task 7* (p29), *Link Task 11* (p29), *Link Task 14* (p30) and *Link Task 39* (p49). There was one difference on the questions of the *Mathematics Survey* (p55). Boys gave a more positive response than girls to the question, “How do you feel about doing things in maths you haven’t tried before?”

### Ethnicity

Results achieved by Māori, Pasifika, and Pakeha (all other) students were compared using effect-size procedures. First, the results for Pakeha students were compared to those for Māori students. Second, the results for Pakeha students were compared to those for Pasifika students.

### Pakeha-Māori Comparisons

For year 4 students, the mean-effect size across the 63 tasks was .37 (Pakeha students averaged 0.37 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences ( $p < .01$ ) on 41 of the 63 tasks. Pakeha students scored higher than Māori students on all 41 tasks. Because of the number of tasks showing differences, they are not listed here. There was one difference on questions of the *Mathematics Survey* (p55). Māori students were more positive than Pakeha students in response to the question, “How much do you like doing maths at school?”

For year 8 students, the results were similar. The mean-effect size across the 89 tasks was .35 (Pakeha students averaged 0.35 standard deviations

higher than Māori students). This is a moderate difference. There were statistically significant differences on 52 of the 89 tasks. Pakeha students scored higher than Māori students on all 52 tasks. Because of the number of tasks showing differences, they are not listed here. There was one difference on the questions of the *Mathematics Survey* (p55). Māori students were more positive than Pakeha students in response to the question, “How good does your teacher think you are at maths?”

### Pakeha-Pasifika Comparisons

Readers should note that only 31 to 41 Pasifika students were included in the analysis for each task. This is lower than normally preferred for NEMP subgroup analyses, but has been judged adequate for giving a useful indication, through the overall pattern of results, of the Pasifika students’ performance. Because of the relatively small numbers of Pasifika students,  $p = .05$  has been used here as the critical level for statistical significance.

For year 4 students, the mean-effect size across the 63 tasks was .35 (Pakeha students averaged 0.35 standard deviations higher than Pasifika students). This is a moderate difference. There were statistically significant differences on 25 of the 63 tasks. Pakeha students scored higher on all 25 tasks. Because of the number of tasks showing differences, they are not listed here. There were also differences on four questions of the *Mathematics Survey* (p55). Pasifika students were more positive than Pakeha students in response to the questions, “How good do you think you are at maths?” “How much do you like doing maths with others?”, “How much do you like helping others with their maths?” and “How do you feel about learning or doing maths as you get older?”

For year 8 students, the mean-effect size across the 89 tasks was .51 (Pakeha students averaged 0.51 standard deviations higher than Pasifika students). This is a large difference. There were statistically significant differences on 60 of the 89 tasks. Pakeha students scored higher on all 60 tasks. Because of the number of tasks showing differences, they are not listed here. There were

no differences on questions of the *Mathematics Survey* (p55).

### Home Language

Results achieved by students who reported that English was the predominant language spoken at home were compared, using effect-size procedures, with the results of students who reported predominant use of another language at home (most commonly an Asian or Pasifika language). Because of the relatively small numbers in the “other language” group,  $p = .05$  has been used here as the critical level for statistical significance.

For year 4 students, the mean-effect size across the 63 tasks was 0.10 (students for whom English was the predominant language at home averaged 0.10 standard deviations higher than the other students). This is a small difference. There were statistically significant differences on five of the 63 tasks: *Maths Helper* (p15), *Torn Tape* (p40), *Trapezium* (p45), *Pick a Teddy* (p51) and *Link Task 29* (p42). For each of these five tasks, the students for whom English was the predominant language at home performed significantly better than the students who reported using another language at home. There were statistically significant differences on seven questions of the *Mathematics Survey* (p55): “How much do you like doing maths at school?”, “Would you like to do more, the same or less maths at school?”, “How much do you like doing maths on your own?”, “How much do you like helping others with their maths?”, “How do you feel about doing things in maths you haven’t tried before?”, “How much do you like doing maths in your own time?” and “How do you feel about learning or doing maths as you get older?” The students who reported using another



language at home were more positive than the students for whom English was the predominant language at home on all seven questions.

For year 8 students, the mean-effect size across the 89 tasks was .10 (students for whom English was the predominant language at home averaged 0.10 standard deviations higher than the other students). This is a small difference. There were statistically significant differences on nine of the 89 tasks. Students for whom English was the predominant language spoken at home scored higher on eight of these tasks: *Maths Helper* (p15), *Show Me The Time* (p33), *Torn Tape* (p40), *Nets* (p46),

### Summary, with Comparisons to Previous Mathematics Assessments

Community size, school size, school type (full primary, intermediate, or year 7 to 13 high school), and geographic zone were not important factors predicting achievement on the mathematics tasks. The same was true for the 2001 and 1997 assessments. However, there were statistically significant differences in the performance of students from low, medium and high decile schools on 62.5 percent of the tasks at year 4 level (compared to 87 percent in 2001 and 85 percent in 1997), and 65 percent of the tasks at year 8 level (compared to 76 percent in 2001 and 77 percent in 1997). The change for year 4 students is noteworthy.

For the comparisons of boys with girls, Pakeha with Māori, Pakeha with Pasifika students, and students for whom the predominant language at home was English with those for whom it was not, effect sizes were used. Effect size is the difference in mean (average) performance of the two groups, divided by the pooled standard deviation of the scores on the particular task. For this summary, these effect sizes were averaged across all tasks.

Year 4 boys averaged slightly higher than girls, with a mean effect size of 0.08 (very similar to the mean effect size of 0.10 in 2001). Year 8 girls averaged slightly higher than boys, with a mean effect size of 0.03 (the same as in 2001). As was also true in 2001, the mathematics survey results at both year levels showed some evidence that boys were more positive than girls about mathematics activities.

*Chocolate Bars* (p52), *Link Task 29* (p42), *Link Task 34* (p42) and *Link Task 47* (p52). Students who reported using a language other than English at home scored higher on *Flies at the Barbecue* (p22). There were also differences on three questions of the *Mathematics Survey* (p55): “How much do you like doing maths in your own time?”, “How much do you like helping others with their maths?” and “How do you feel about learning or doing maths as you get older?” The students who reported using another language at home were more positive than the students for whom English was the predominant language at home on all three questions.

Pakeha students averaged moderately higher than Māori students, with mean effect sizes of 0.37 for year 4 students and 0.35 for year 8 students (the corresponding figures in 2001 were 0.46 and 0.42). The responses to the questions of the mathematics survey yielded only one difference at each year level.

Year 4 Pakeha students averaged moderately higher than Pasifika students, with a mean effect size of 0.35 (compared to 0.59 in 2001). This is a noteworthy change. Year 8 Pakeha students also averaged substantially higher than Pasifika students, with a mean effect size of 0.51 (compared to 0.53 in 2001). The responses to the *Mathematics Survey* (p55) showed some differences at year 4, with the Pasifika students indicating more positive responses than the Pakeha students.

Compared to students for whom the predominant language at home was English, students from homes where other languages predominated averaged slightly lower, with mean effect sizes of 0.10 for year 4 students and 0.10 for year 8 students. Comparative figures are not available for the assessments in 2001. Year 4 students who reported speaking a language other than English at home were generally more positive about mathematics than students whose predominant language at home was English. These differences largely subsided at year 8.

