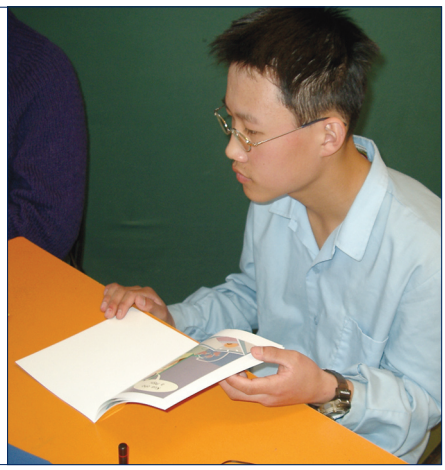
8 Performance of Subgroups

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 an overall score for each task, created by adding together scores for appropriate components of the task.



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 (for *School Type*), 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 three of the five school variables, statistically significant differences among the subgroups were found for less than 10 percent of the tasks at both year 4 and year 8. For the remaining two variables, statistically significant differences were found on higher proportions of tasks. In the detailed report below, all "differences" mentioned are statistically significant (to save space, the words "statistically significant" are omitted).

School Type

Results were compared for year 8 students attending full primary schools and year 8 students attending intermediate schools. There were differences between the two subgroups on one of the 31 reading tasks and one of the 21 speaking tasks. Students from intermediate schools scored higher on *Link Task 2* (p26) and lower on *Thank You* (p56). There were no differences on questions of the year 8 *Reading and Speaking Survey (p62)*.

School Size

Results were compared from students in large, medium sized, and small schools (exact definitions were given in Chapter 1).

For year 4 students, there were differences among the subgroups on two of the 25 reading tasks, with students from large schools scoring highest and students from small schools scoring lowest on two oral reading tasks, *Reading Record – Fiction* (p17) and *Reading Record – Non-Book* (p20). There were no differences on any of the 20 speaking tasks, or on questions of the year 4 *Reading and Speaking Survey* (p61).

For year 8 students, there were differences among the subgroups on two of the 31 reading tasks, with students from large schools scoring high on *Link Task 2* (p26) and *Toyworld* (p30). There were also differences on two of the 21 speaking tasks, with students from large schools scoring highest on *Link Task 16* (p47) and students from small schools scoring highest on *Thank You* (p56). There were no differences on questions of the year 8 *Reading and Speaking Survey* (p62).

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 one of the 25 reading tasks, with students from Auckland scoring highest on *What a Ride!* (p35). There were no differences on any of the 20 speaking tasks or on questions of the year 4 *Reading and Speaking Survey* (p61).

For year 8 students, there were differences among the three subgroups on three of the 31 reading tasks. Students from the South Island scored distinctly lowest and students from the North Island beyond Auckland scored highest on reading of Māori text in *Link Task 1* (p26). Students from the South Island scored highest, and students from Auckland lowest on *Wallabies in New Zealand* (p28), and students from Auckland scored lowest on *Link Task 7* (p41). There were no differences on any of the 21 speaking tasks or on questions of the year 8 *Reading and Speaking Survey* (p62).

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 fewer than 10,000 people (rural areas).

For year 4 students, there were differences among the three subgroups on seven of the 25 reading tasks. Students from the main centres scored highest on *Reading Record – Fiction* (p17), *Reading Record – Non-Book* (p20), *Nonsense You Might See and Hear* (p24), and *Link Task 5* (p41). Students from rural areas scored lowest on the first two of those tasks and also on *Link Task 3* (p26), *Toyworld* (p30), and *What a Ride!* (p35). There were also differences on two of the 20 speaking tasks, with students from rural areas scoring highest on *Hāngi* (p43), but lowest (with students from provincial cities highest) on *Thank You* (p56). There were no differences on questions of the year 4 *Reading and Speaking Survey* (p61).

For year 8 students, there were no differences among the three subgroups on any of the 31 reading tasks, but there were differences on two of the 21 speaking tasks. Students from provincial cities scored highest on *Thank You* (p56), while students from rural areas scored lowest on *Link Task 17* (p57). There were no differences on questions of the year 8 *Reading and Speaking Survey* (p62).

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 SES index takes into account household income levels, categories of employment and the ethnic mix in the census mesh blocks. The SES index uses 10 subdivisions, each containing ten percent of schools (deciles 1 to 10). For our purposes, the bottom three deciles (1-3) formed the low SES group, the middle four deciles (4-7) formed the medium SES group, and the top three deciles (8-10) formed the high SES group. Results were compared for students attending schools in each of these three SES groups.

For year 4 students, there were differences among the three subgroups on 20 of the 25 reading tasks and 12 of the 20 speaking tasks. In each case, students from low SES schools scored lowest. Students from low decile schools scored highest on one task involving reading Māori text, *Link Task 1* (p26). On the remaining 19 reading tasks and the 12 speaking tasks, students from low decile schools scored lower than students from high decile schools. Because of the large number of tasks, they are not listed here. There were no differences on questions of the year 4 *Reading and Speaking Survey* (p61).

For year 8 students, there were differences among the three subgroups on 26 of the 31 reading tasks and 11 of the 21 speaking tasks. Students from low decile schools scored highest on one task involving reading Māori text, *Link Task 1* (p26). On the remaining 25 reading tasks and 11 speaking tasks, students from low decile schools scored lowest. While students from high SES schools generally did better than students from medium SES schools, these differences were usually smaller than the performance differences between students from low and medium SES schools. Because of the large number of tasks, they are not listed here. There were no differences on questions of the year 8 *Reading and Speaking Survey* (p62).

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 previous cycle of the Project (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 four comparisons, differences in task performance between the two subgroups are described using "effect sizes" and statistical significance.

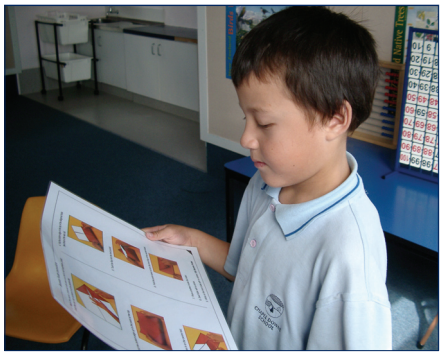
For each task and each year level, the analyses began with a t-test comparing the performance of the two selected subgroups and checking for statistical significance of the differences. Then the mean score obtained by students in one subgroup was subtracted from the mean score obtained by students in 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 the effect size procedures.

For year 4 students, the mean effect size across the 25 reading tasks was -.20 (girls averaged 0.20 standard deviations higher than boys). This is a small difference. There were statistically significant ($p < .01$) differences favouring girls on 12 of the 25 tasks: three involving oral reading and nine involving comprehension. The mean effect size across the 19 speaking tasks was -.07 (girls averaged 0.07 standard deviations higher than boys). This is a small difference. There was a statistically significant difference favouring girls on just one speaking task: *Link Task 21* (p57). Girls also gave more positive ratings than boys on eight questions of the year 4 *Reading and Speaking Survey* (p61). They reported greater enjoyment of reading at school (question 1) and in their own time (question 8), and were more positive about receiving a book as a present (question 9), the stories or books in their



school reading programme (question 12), how well they thought they read (question 14), reading out loud to the teacher (question 16), reading out loud to the class (question 17) and talking to the whole class (question 18).

For year 8 students, the mean effect size across the 31 reading tasks was -.15 (girls averaged 0.15 standard deviations higher than boys). This is a small difference. There were statistically significant ($p < .01$) differences favouring girls on seven of the 31 tasks: two involving oral reading and five involving comprehension. The mean effect size across the 20 speaking tasks was -.07 (girls averaged 0.07 standard deviations higher than boys). This is a small difference. There was a statistically significant difference favouring girls on three speaking tasks: *Lively Poems* (p52), *Link Task 21* (p57) and *Link Task 22* (p57). Girls also were more positive than boys on just one question of the year 8 *Reading and Speaking Survey* (p62): their enjoyment of the stories or books in their school reading programme (question 12).

Ethnicity

Results achieved by Māori, Pasifika and Pakeha (all other) students were compared using the 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

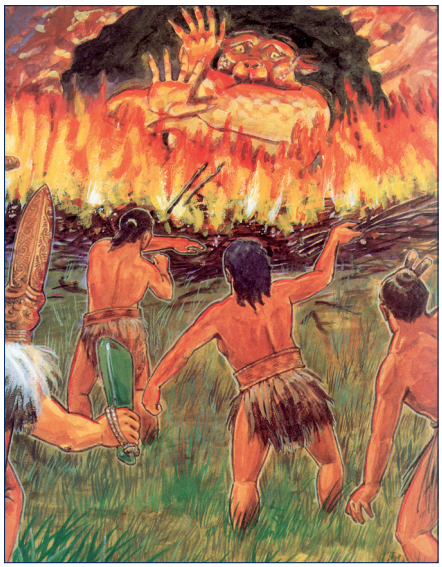
For year 4 students, the mean effect size across the 25 reading tasks was +.37 (Pakeha students averaged 0.37 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant ($p < .01$) differences on 17 of the 25 tasks, with Pakeha students higher on all of these tasks. The mean effect size across the 19 speaking tasks was +.29 (Pakeha students averaged 0.29 standard deviations higher than Māori students). This is a moderate difference. Pakeha students scored statistically significantly higher on nine of the 19 tasks. There was also a statistically significant difference on one question of the year 4 *Reading and Speaking Survey* (p61): Māori students reported more guidance from the teacher to help them improve their reading (question 6).

For year 8 students, the mean effect size across the 31 reading tasks was +.31 (Pakeha students averaged 0.31 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences on 20 of the 31 tasks, with Pakeha students higher on 18 tasks involving reading in English and Māori students higher on the two tasks involving reading in Māori. The mean effect size across the 20 speaking tasks was +.20 (Pakeha students averaged 0.20 standard deviations higher than Māori students). This is a small difference. Pakeha students scored statistically significantly higher on six of the 20 tasks. There were also statistically significant differences on three questions of the year 8 *Reading and Speaking Survey* (p62): Māori students reported more guidance from the teacher to help them improve their reading (question 6), were less enthusiastic about reading in their own time (question 8) and were less enthusiastic about receiving a book as a present (question 9).

Pakeha-Pasifika Comparisons

Readers should note that only 20 to 50 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 25 reading tasks was +.27 (Pakeha students averaged 0.27 standard deviations higher than Pasifika students). This is a moderate difference. There were statistically significant differences on seven of the 25 tasks, with Pakeha students higher on all of these tasks, which involved comprehension rather than just oral reading. The mean effect size across the 19 speaking tasks was +.35 (Pakeha students averaged 0.35 standard deviations higher than Pasifika students). This is a moderate difference. Pakeha students scored statistically significantly higher on seven of the 19 tasks. There were also statistically significant differences on two questions of the year 4 *Reading and Speaking Survey* (p61): Pasifika students were more positive about reading at school (question 1) and reported fewer opportunities to talk to others in their class (question 21).



For year 8 students, the mean effect size across the 31 reading tasks was +.41 (Pakeha students averaged 0.41 standard deviations higher than Pasifika students). This is a large difference. There were statistically significant differences on 24 of the 31 tasks, with Pakeha students higher on 23 tasks involving reading in English and Pasifika students higher on one task involving reading in Māori. The mean effect size across the 20 speaking tasks was +.27 (Pakeha students averaged 0.27 standard deviations higher than Pasifika students). This is a moderate difference. Pakeha students scored statistically significantly higher on six of the 20 tasks. There were also statistically significant differences on two questions of the year 8 *Reading and Speaking Survey* (p62): Pasifika students reported more guidance from the teacher to help them improve their reading (question 6) and more opportunities to read to others at school (question 7).

Home Language

Results achieved by students who reported that English was the predominant language spoken at home were compared, using the 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 25 reading tasks was +.21 (students for whom English was the predominant language at home averaged 0.21 standard deviations higher than the other students). This is a moderate difference. There were statistically significant differences on 13 of the 25 tasks: students for whom English was the predominant language spoken at home scored higher on 12 tasks but lower on one task involving reading Māori text. The mean effect size across the 19 speaking tasks was +.17 (students for whom English was the predominant language at home averaged 0.17 standard deviations higher than the other students). This is a small difference. There were statistically significant differences, favouring those for whom English was the predominant language spoken at home, on two tasks: *Lively Poems* (p52), and *Link Task 22* (p57). There were also statistically significant differences on two questions of the year 4 *Reading and Speaking Survey* (p61): students for whom the predominant language at home was not English reported less positive parental views about their reading (question 4) but were more positive about reading out loud in class (question 17).

For year 8 students, the mean effect size across the 31 reading tasks was +.16 (students for whom English was the predominant language at home averaged 0.16 standard deviations higher than the other students). This is a small difference. There were statistically significant differences on nine of the 31 tasks: students for whom English was the predominant language spoken at home scored higher on eight tasks but lower on one task involving reading Māori text. The mean effect size across the 20 speaking tasks was +.14 (students for whom English was the predominant language at home averaged 0.14 standard deviations higher than the other students). This is a small difference. There was a statistically significant difference, favouring those for whom English was the predominant language spoken at home, on just one task: *Link Task 16* (p47). There were also statistically significant differences on seven questions of the year 8 *Reading and Speaking Survey* (p62): students for whom the predominant language at home was not English reported less positive parental views about their reading (question 4) and fewer opportunities to talk to others in class (question 21), but were more positive about reading in their own time (question 8), receiving a book as a present (question 9), looking at books in a bookshop (question 10), going to a library (question 11) and the stories or books in their reading programme at school (question 12).

Summary, With Comparisons to Previous Reading and Speaking Assessments

School type (full primary or intermediate), school size, community size and geographic zone did not seem to be important factors predicting achievement on the reading and speaking tasks. The same was true for the 2000 and 1996 assessments. However, for year 4 students there were statistically significant differences in the performance of students from low, medium and high decile schools on 80 percent of the reading tasks (compared to 88 percent in 2000 and 71 percent in 1996), and 60 percent of the speaking tasks (compared to 87 percent in 2000 and 75 percent in 1996). There were also differences for year 8 students on 84 percent of the reading tasks (which compares with 58 percent in 2000 and 93 percent in 1996), and 52 percent of the speaking tasks (which compares with 56 percent in 2000 and 67 percent in 1996).

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 tasks.

Girls averaged higher than boys on reading tasks, with mean effect sizes of 0.20 for year 4 students and 0.15 for year 8 students (the corresponding figures in 2000 were 0.25 and 0.10). These differences in 2004 are small, with little change since 2000. The reading survey results showed that year 4 girls were markedly more enthusiastic about reading than year 4 boys, but there was little difference between year 8 girls and boys. On speaking tasks, the advantage of girls over boys was very small, with mean effect sizes of 0.07 at both year levels (the corresponding figures in 2000 were 0.24 and 0.06). This indicates a substantial reduction in disparity for year 4 students, with no change in the already very small disparity for year 8 students.

Pakeha students averaged higher than Māori students on the tasks involving reading in English, with a large mean effect size of 0.41 for year 4 students and a moderate effect size of 0.36 for year 8 students (the corresponding figures in 2000 were 0.63 and 0.35). This indicates a substantial reduction in disparity for year 4 students, but no change for year 8 students. Māori students averaged higher than Pakeha students on the two tasks involving reading in Māori, with a small mean effect size of 0.05 for year 4 students and a large mean effect size of 0.53 for year 8 students (the corresponding figures in 2000 were 0.35 and 0.79). Pakeha students performed better than Māori students on speaking tasks, with moderate mean effect sizes of 0.29 for year 4 students and 0.20 for year 8 students (the corresponding figures in 2000 were 0.41 for year 4 students and 0.35 for year 8 students). This indicates a moderate reduction in disparity at both year levels.

Pakeha students averaged higher than Pasifika students on the tasks involving reading in English, with a moderate mean effect size of 0.31 for year 4 students and a large mean effect size of 0.48 for year 8 students (the corresponding figures in 2000 were 0.64 and 0.60). This indicates a substantial reduction in disparity for year 4 students, with a small reduction for year 8 students. Pasifika students averaged higher than Pakeha students on the two tasks involving reading in Māori, with a small mean effect size of 0.15 for year 4 students and a large mean effect size of 0.50 for year 8 students (the corresponding figures in 2000 were 0.47 and a very large 1.12). Pakeha students performed better than Pasifika students on speaking tasks, with moderate mean effect sizes of 0.35 for year 4 students and .27 for year 8 students (the corresponding figures in 2000 were 0.77 and 0.47, but these were based on a more restricted range of tasks).

Compared to students for whom the predominant language spoken at home was not English, students for whom the predominant language at home was English scored higher at both year levels on tasks involving reading and speaking in English. For reading in English, there was a moderate mean effect size of 0.26 for year 4 students and a small mean effect size of 0.18 for year 8 students. The corresponding figures for speaking tasks were 0.17 and 0.14, both small mean effect sizes. The students for whom the predominant language at home was not English scored higher at both year levels on the two tasks involving reading in Māori, with moderate mean effect sizes of 0.35 for year 4 students and 0.26 for year 8 students. No corresponding effect sizes from 2000 are available for any of these comparisons.