5 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 two or three subgroups on each variable, as detailed in Chapter 1 (p7).

The analyses of the relative performance of subgroups used a total 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 (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 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 the first four of the five school variables, statistically significant differences among the subgroups were found for less than 12 percent of the tasks at both year levels. For the remaining variable, statistically significant differences were found on more than half 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).

The performance patterns found were different for the listening tasks (Chapter 3) and the viewing tasks (Chapter 4), so the results are discussed separately for these two strands of the English curriculum.

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 no differences among the subgroups on any of the 17 listening tasks, nor any of the 14 viewing tasks. For year 8 students, there were no differences on any of the 20 listening tasks, nor any of the 16 viewing tasks.

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 towns) and communities containing less than 10,000 people (rural areas).

For year 4 students, there were no differences among the subgroups on any of the 17 listening tasks, nor any of the 14 viewing tasks. For year 8 students, there were no differences on any of the 20 listening tasks, nor any of the 16 viewing tasks.

School Type

Results were compared for year 8 students attending full primary and intermediate schools. There were no differences between these two sub-groups on any of the 16 viewing tasks, but there was a difference on one of

the 20 listening tasks. Students from full primary schools scored higher than students from intermediate schools on *Link Task 1* (p27)

There are now enough year 8 students attending year 7 to 13 high schools to permit comparisons between them and the students attending intermediate schools. There were no differences between these two sub-groups on any of the 16 viewing tasks, but there was a difference on one of the 20 listening tasks. Students from year 7 to 13 high schools scored higher than students from intermediate schools on *Link Task 1* (p27).

Zone

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

For year 4 students, there was a difference on one of the 14 viewing tasks. On *Link Task 15* (p41), Auckland students had the highest scores, followed by South Island students, and then the remaining North Island students. There was also a difference on one of the 17 listening tasks: Auckland students scored lower than students from the South Island or the rest of the North Island on *Colour Cat* (p19).

For year 8 students, there were no differences on any of the 16 viewing tasks, but there were differences on two of the 17 listening tasks. Students from Auckland scored lowest on *Colour Cat* (p19) and *Link Task 4* (p27).

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 ten 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 15 of the 17 listening tasks and eight of the 14 viewing tasks. Because of the large number of tasks involved, they will not be listed here. In all cases, students in the low SES schools scored lowest. While students from high SES schools generally did better than students from medium SES school, these differences were often smaller than the differences between students from low and medium SES schools.

For year 8 students, there were differences among the three subgroups on 15 of the 20 listening tasks and 11 of the 16 viewing tasks. For about half of these tasks, the prominent feature was the low performances of students in the low SES schools, with more modest differences between students from medium and high SES schools. For the remaining tasks showing differences, the performance gaps were evenly distributed from low through to high SES schools.

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.

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 +0.30, 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 procedure.

For year 4 students, the mean effect size across 17 listening tasks was 0.09 (girls averaged 0.09 standard deviations higher than boys). This is a small difference. There were statistically significant differences on two of the 17 tasks. Girls scored higher than boys on *New Student* (p17) and *Link Task 6* (p27). The mean effect size across 14 viewing tasks was 0.02 (girls

averaged 0.02 standard deviations higher than boys). This is a negligible difference. There were no statistically significant differences on any of the 14 viewing tasks.

For year 8 students, the mean effect size across 20 listening tasks was 0.10 (girls averaged 0.10 standard deviations higher than boys). This is a small difference. There were statistically significant differences on six of the 20 tasks. Girls scored higher than boys on *New Student* (p17), *Colour Cat* (p19), and *Link Tasks 6, 7* and *8* (p27). Boys scored higher than girls on *Link Task 2* (p27). The mean effect size across 16 viewing tasks was 0.09 (girls averaged 0.09 standard deviations higher than boys). This is a small difference. There were no statistically significant differences on any of the 16 viewing tasks.

Student Ethnicity

Two sets of comparisons were made by ethnic groups. First, Pakeha students were compared to Māori students, and then Pakeha students were compared to Pasifika students. It should be noted that "Pakeha students" includes all students not classified as Maori or Pasifika. Because of the relatively small number of Pasifika students includes in the analysis for each task, a statistical significance level of 0.05 was used for determining differences in the Pakeha/Pasifika comparisons. The Pakeha/Māori comparisons used 0.01 as the statistical significance level.

Pakeha/Māori Comparisons

For year 4 students, the mean effect size across 17 listening tasks was 0.47 (Pakeha students averaged 0.47 standard deviations higher than Māori students). This is a large difference. There were statistically significant differences favouring Pakeha students on 13 of the 17 listening tasks. Because of the large number of tasks, they are not listed here. The mean effect size across 14 viewing tasks was 0.29 (Pakeha students averaged 0.29 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences favouring Pakeha students on four of the 14 viewing tasks: *Silent Ads* (p30), *Storm-Fish* (p35), and *Link Tasks 11* and *14* (p41).

For year 8 students, the mean effect size across 20 listening tasks was 0.33 (Pakeha students averaged 0.33 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences favouring Pakeha students on 7 of the 20 listening tasks: *School Notices* (p16), *Leonard King* (p20), *Little Poems* (p23), *Butterfly or Moth* (p24), and *Link Tasks 1, 3* and *9* (p27). The mean effect size across 16 viewing tasks was 0.30 (Pakeha students averaged 0.30 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences favouring Pakeha students on four of the 16 viewing tasks: *Book Cover* (p29), *Marmite* (p37), *Breakfast Foods* (p38), and *Link Task 10* (p41).

Pakeha/Pasifika Comparisons

For year 4 students, the mean effect size across 17 listening tasks was 0.55 (Pakeha students averaged 0.55 standard deviations higher than Pasifika students). This is a large difference. There were statistically significant differences on 14 of the 17 listening tasks. Because of the large number of tasks, they are not listed here. The mean effect size across 14 viewing tasks was 0.26 (Pakeha students averaged 0.26 standard deviations higher than Pasifika students). This is a moderate difference. There were statistically significant differences favouring Pakeha students on three of the 14 viewing tasks: *Breakfast Foods* (p38) and *Link Tasks 11* and *15* (p41).

For year 8 students, the mean effect size across 20 listening tasks was 0.61 (Pakeha students averaged 0.61 standard deviations higher than Pasifika students). This is a large difference. There were statistically significant differences favouring Pakeha students on 18 of the 20 listening tasks. Because of the large number of tasks, they are not listed here. The mean effect size across 16 viewing tasks was 0.40 (Pakeha students averaged 0.40 standard deviations higher than Pasifika students). This is a large difference. There were statistically significant differences favouring Pakeha students on seven of the 16 viewing tasks: *Silent Ads* (p30), *Giant Weta and Giant Dragonfly* (p34), *Storm-Fish* (p35), *Breakfast Foods* (p38), and *Link Tasks 12, 14* and *15* (p41).

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). A statistical significance level of 0.05 is used because of quite small numbers of children in the "other language" group.

For year 4 students, the mean effect size across 17 listening tasks was 0.24 (students for whom English was the predominant language at home averaged 0.24 standard deviations higher than the other students). This is a moderate difference. There were statistically significant differences favouring students whose home language was English on eight of the 17 listening tasks: *Possums* (p15), *Porridge* (p18), *Colour Cat* (p19), *Butterfly or Moth?* (p24), *Zak* (p25), and *Link Tasks 2, 3,* and *5* (p27). The mean effect size across 14 viewing tasks was 0.14 (year 4 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 were statistically significant differences favouring students whose home language was English on two of the 14 viewing tasks: *Book Cover* (p29) and *Breakfast Foods* (p38).

For year 8 students, the mean effect size across 20 listening tasks was 0.28 (students for whom English was the predominant language at home averaged 0.28 standard deviations higher than the other students). This is a moderate difference. There were statistically significant differences on eight of the 20 listening tasks: *Possums* (p15), *Porridge* (p18), *Colour Cat* (p19), *Leonard King* (p20), and *Link Tasks 2, 3, 7* and *9* (p27). The mean effect size across 16 viewing tasks was 0.14 (year 8 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 were statistically significant differences favouring students whose home language was English on two of the 16 viewing tasks: *Tick Tick* (p40) and *Link Task 12* (p41).

Summary, with Comparisons to Previous Listening and Viewing Assessments

School size, school type (full primary, intermediate, or year 7 to 13 high school) and community size were not important factors predicting achievement on the listening and viewing tasks. These results parallel those from the 2002 and 1998 assessments.

There were differences by zone (region) for less than 12 percent of the listening and viewing tasks at both year levels. At year 4 level only, this represents a change from the 2002 assessments, where South Island students scored higher than Auckland students on 36 percent of listening tasks and 44 percent of viewing tasks. The results from 2006 are similar to the 1998 results, which saw few differences by zone at both year levels.

There were statistically significant differences in the performance of students from low, medium and high decile schools on 88 percent of the listening tasks at year 4 level (compared to 71 percent in 2002 and 87 percent in 1998), and 75 percent of the listening tasks at year 8 level (compared to 59 percent in 2002 and 78 percent in 1998). Overall, there has been little reduction in disparities of achievement on listening tasks between 1998 and 2006. For the viewing tasks, there were differences on 57 percent of the tasks at year 4 level (compared to 50 percent in 2002 and 100 percent in 1998), and 69 percent of the tasks at year 8 level (compared to 61 percent in 2002 and 86 percent in 1998). The reductions in disparities of achievement on viewing tasks observed between 1998 and 2002 have been maintained in 2006.

Girls averaged slightly higher than boys on listening tasks at both year levels, with a mean effect size at year 4 level of 0.09 (slightly reduced from 0.13 in 2002) and a mean effect size at year 8 level of 0.10 (reduced from 0.19 in 2002). On the viewing tasks, gender differences also favoured girls but were small at both year levels, both in 2006 and earlier in 2002. The mean effect size at year 4 was 0.02 (slightly reduced from 0.05 in 2002), while at year 8 level it was 0.09 (slightly increased from 0.06 in 2002).

Pakeha students averaged higher than Māori students on the listening tasks, with a large mean effect size of 0.47 for year 4 students (increased from 0.34 in 2002), and a moderate mean effect size of 0.33 for year 8 students (little changed from 0.29 in 2002). On the viewing tasks, Pakeha students scored moderately higher than Māori students at both year levels. The mean effect size for year 4 students was 0.29 (little changed from 0.32 in 2002), while for year 8 students the mean effect size was 0.30 (little changed from 0.31 in 2002).

Pakeha students averaged substantially higher than Pasifika students on the listening tasks, with a large mean effect size of 0.55 for year 4 students (reduced from 0.71 in 2002), and a similarly large mean effect size of 0.61 for year 8 students (little changed from 0.63 in 2002). On the viewing tasks, Pakeha students scored moderately higher than Pasifika students at year 4 level and more strongly higher at year 8 level. The mean effect size for year 4 students was 0.26 (substantially reduced from 0.43 in 2002), while for year 8 students the mean effect size was 0.40 (reduced from 0.51 in 2002).

Compared to students for whom the predominant language at home was not English, students from homes where English predominated averaged moderately higher on listening tasks (mean effect sizes 0.24 at year 4 level and 0.28 at year 8 level). For viewing tasks, the advantage for students from homes where English predominated was smaller, with small mean effect sizes of 0.14 at both year levels. Comparative effect sizes are not available from the 2002 assessments.