Performance of Subgroups 7

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 (p8).

Analyses of the relative performance of subgroups used the total score for each task, created as described in Chapter 1 (p8).

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 two of the five school variables, statistically significant differences among the subgroups were found for less than seven percent of the tasks at both year levels For the next two variables, statistically significant differences were found for less than seven percent at year 8 level, but 20 to 30 percent of the tasks at year 4 level. 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).

School Type

Results were compared for year 8 students attending full primary and intermediate (or middle) schools. There were no differences between these two subgroups on any of the 33 tasks, or on questions of the year 8 *Writing Survey* (p58).

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 statistically significant differences (p<.01) on two of the 33 tasks. Students from year 7 to 13 high schools scored higher on *Link Task 2* (p34) and *Torch* (p38). There was also a difference on one question of the year 8 *Writing Survey* (p58), with students from intermediate schools indicating that teachers read their work more often (question 9).

School Size

Results were compared from students in large, medium-sized, and small schools. Exact definitions were given in Chapter 1 (p8).

For year 4 students, there were differences among the three subgroups on two of the 30 tasks: *Link Task 4* (p34) and *Spelling List* (p52). On both of these tasks, students from small schools scored lowest and students from large schools highest. There were no differences on any questions of the year 4 *Writing Survey* (p57).

For year 8 students, there was a difference on just one of the 33 tasks, with students from small schools scoring lowest (and students from large schools highest) on *Link Task 1* (p34). There were no differences on questions of the year 8 *Writing Survey* (p58).

Community Size

Results were compared for students living in communities containing over 100,000 people (main centre), communities containing 10,000 to 100,000 people (provincial city) 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 30 tasks. Students from rural areas scored lowest on all six tasks: *For or Against?* (p22), *Link Task 1* (p34), *After School* (p40), *Link Task 7* (p48), *Parts of Speech* (p50) and *Spelling List* (p52). There were no differences on questions of the year 4 *Writing Survey* (p57).

For year 8 students, there were no differences on any of the 33 tasks. There was, however, a difference on one question of the year 8 *Writing Survey* (p58), with students from main centres more positive about writing at school (question 1).

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 30 tasks. Students from Auckland scored clearly highest on five tasks: *Link Task 4* (p34), *Jenny's Letter* (p47), *Link Tasks 8* and *10* (p48) and

Parts of Speech (p50). Students from other parts of the North Island scored clearly lowest on *Popcorn* (p43), *Shells* (p46) and *Link Task 14* (p54). Students from the South Island scored clearly highest on *Link Task 6* (p48). There was also a difference on one question of the year 4 *Writing Survey* (p57): students from large schools indicated that they least often had "others" read what they wrote (question 13).

For year 8 students, there were differences among the three subgroups on two of the 33 tasks: students from Auckland scored lowest on *A Day I'll Never Forget* (p16), but highest on *After School* (p40). There was also a difference on one question of the year 8 *Writing Survey* (p58), with students from Auckland most positive about writing in their own time (question 5).

Socio-Economic Index (SES)

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 and categories of employment. The SES index 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 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 19 of the 30 tasks, six or seven in each of the three task chapters. Because of the number of tasks showing differences, they are not listed here. Students in high decile schools performed better than students in low decile schools on all 19 tasks. There were also differences on six questions of the year 4 *Writing Survey* (p57). Students from low decile schools were most positive about writing in school (question 1) and in their own time (question 5), thought that they spent more time in school writing things like stories, poems or letters (question 7), reported that siblings or "others" read their work more often (questions 11 and 13) and reported more frequent use of writing using a computer at school (question 14).

For year 8 students, there were differences among the three subgroups on 17 of the 33 tasks: five in Chapter 3, eight in Chapter 4 and four in Chapter 5. Because of the number of tasks showing differences, they are not listed here. Students in high decile schools performed better than students in low decile schools on all 17 tasks. There were also differences on four questions of the year 8 *Writing Survey* (p58). Students from low decile schools thought that they spent more time in school writing things like stories, poems or letters (question 7), reported that siblings or friends read their work more often (questions 11 and 12) and reported less frequent use of writing using a computer at home (question 15).

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 +.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 30 tasks was 0.28 (girls averaged 0.28 standard deviations higher than boys). This is a moderate difference. There were statistically significant ($p < .01$) differences favouring girls on 18 of the 30 tasks: six in Chapter 3, eight in Chapter 4 and four in Chapter 5. Because of the number of tasks showing differences, they are not listed here. There were also differences on five questions of the year 4 *Writing Survey* (p57). Girls were more positive about writing in school (question 1) and in their own time (question 5), about how good they thought they were in writing (question 2) and reported that friends or "others" read their work more often (questions 12 and 13).

For year 8 students, the mean-effect size across the 33 tasks was 0.33 (girls averaged 0.33 standard deviations higher than boys): a moderate difference. There were statistically significant differences favouring girls on 24 of the 33 tasks: nine in Chapter 3, eleven in Chapter 4, and four in Chapter 5. Because of the number of tasks showing differences, they are not listed here. There were also differences on eight questions of the year 8 *Writing Survey* (p58). Girls were more positive about writing in school (question 1) and in their own time (question 5), about how good they thought they were in writing (question 2) and spelling (question 6), and about how good their teacher thought they were in writing (question 3). They also reported reading their writing to others more often (question 8) and that friends or "others" read their work more often (questions 12 and 13).

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 30 tasks was 0.34 (Pakeha students averaged 0.34 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences ($p < .01$) on 20 of the 30 tasks, spread across the three task chapters, but with the highest proportion in Chapter 3 (expressive writing). Pakeha students scored higher than Māori students on all 20 tasks. Because of the number of tasks showing differences, they are not listed here. There was a difference on one question of the year 4 *Writing Survey* (p57): Māori students reported that they read their work to "others" more often (question 13).

For year 8 students, differences were smaller. The mean-effect size across the 33 tasks was 0.23 (Pakeha students averaged 0.23 standard deviations higher than Māori students): a moderate difference. There were statistically significant differences on eight of the 33 tasks: *For or Against?* (p22), *Link Task 5* (p34), *Jenny's Letter* (p47), *Link Tasks 7* and *10* (p48), *Parts of Speech* (p50), *Punctuation* (p53) and *Link Task 17* (p54). Pakeha students scored higher than Māori students on all eight tasks. There were also differences on two questions of the year 8 *Writing Survey* (p58), with

Māori students indicating that they more often wrote things like stories, poems or letters at school (question 7) but less often wrote using a computer at home (question 15).

Pakeha-Pasifika Comparisons

Readers should note that only 30 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 30 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 on 11 of the 30 tasks, mainly in the areas of expressive writing (Chapter 3) and punctuation (all three punctuation tasks in Chapter 5). Pakeha students scored higher on all 11 tasks. Because of the number of tasks showing differences, they are not listed here. There were also differences on five questions of the year 4 *Writing Survey* (p57). Pasifika students were more positive about writing in school (question 1) and in their own time (question 5), and reported that siblings, friends and "others" read their work more often (questions 11, 12 and 13).

For year 8 students, the mean-effect size across the 33 tasks was 0.29 (Pakeha students averaged 0.29 standard deviations higher than Pasifika students). This is a moderate difference. There were statistically significant differences on 13 of the 33 tasks, spread evenly across the three task chapters. Pakeha students scored higher on all 13 tasks. Because of the number of tasks showing differences, they are not listed here. There were also differences on five questions of the year 8 *Writing Survey* (p58). Pasifika students were more positive about writing in school (question 1) and in their own time (question 5), reported that they more frequently wrote things like stories, poems or letters at school (question 7), and reported that siblings and friends read their work more often (questions 11 and 12).

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 (34 to 58), p = .05 has been used here as the critical level for statistical significance.

For year 4 students, the mean-effect size across the 30 tasks was 0.01 (students for whom English was the predominant language at home averaged 0.01 standard deviations higher than the other students). This is a negligible difference. There were no statistically significant differences on any of the 30 tasks. There were differences on two questions of the year 4 *Writing Survey* (p57). Students for whom the predominant language at home was not English were more positive about writing in their own time (question 5) and reported that parents read their work more often (question10).

For year 8 students, the mean-effect size across the 33 tasks was 0.14 (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 were statistically significant differences on four of the 33 tasks: *A Day I'll Never Forget* (p16), *For or Against?* (p22), *Link Task 4* (p34), and *Link Task 17* (p54). Students for whom English was the predominant language spoken at home scored higher on these four tasks. There was also a difference on one questions of the year 8 *Writing Survey* (p58): students whose predominant language at home was not English reported that their parents were less positive about how good they were at writing (question 4).

Summary, with Comparisons to Previous Writing Assessments

School type (full primary, intermediate, or year 7 to 13 high school), school size, community size and geographic zone were not important factors predicting achievement on the writing tasks at year 8 level. The same was true for the 2002 and 1998 assessments. The evidence was more mixed at year 4 level, where there were statistically significant differences in school size for seven percent of tasks (compared to six percent in 2002 and zero percent in 1998). There were differences by community size for 20 percent of the tasks and by zone (region) for 30 percent of the tasks. Comparative figures for community size and zone from earlier writing assessments were nil percent and 14 percent in 2002, and four percent and 13 percent in 1998.

There were statistically significant differences in the performance of students from low, medium and high decile schools on 63 percent of the tasks at year 4 level (compared to 72 percent in 2002 and 83 percent in 1998) and 52 percent of the tasks at year 8 level (compared to 83 percent in 2002 and 72 percent in 1998). These changes indicate a useful reduction in disparities of achievement.

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 girls averaged moderately higher than boys, with a mean effect size of 0.28 (similar to the effect size of 0.24 in 2002). Year 8 girls also averaged moderately higher than boys, with a mean effect size of 0.34 (reduced a little from 0.40 in 2002). As was also true in 2002, the writing survey results at both year levels showed quite strong evidence that girls were more positive than boys about writing activities.

Pakeha students averaged moderately higher than Māori students, with mean effect sizes of 0.34 for year 4 students and 0.23 for year 8 students (the corresponding figures in 2002 were 0.34 and 0.38, so the 2006 results represent a useful reduction of disparities for year 8 students).

Pakeha students averaged moderately higher than Pasifika students, with mean effect sizes of 0.25 for year 4 students and 0.29 for year 8 students (revealing strongly reduced disparities of performance compared to 2002, when the effect sizes were 0.50 and 0.52). As was also true in 2002, the writing survey results showed that Pasifika students were more enthusiastic about writing and more involved in sharing their writing with others.

Compared to students for whom the predominant language at home was English, students from homes where other languages predominated performed comparably well at year 4 level and slightly lower at year 8 level, with effect sizes of 0.01 and 0.13 respectively. Comparative figures are not available for the assessments in 2002.