

# 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 (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 four of the five school variables, statistically significant differences among the subgroups were found for less than 17 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 movement skills tasks (Chapter 4) and the other tasks (Chapters 3, 5 and 6). In this chapter, the former are referred to as PE (physical education) tasks, the latter as health tasks but it should be noted that physical education involves more than movement skills.

### 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 36 health tasks, on any questions of the year 8 *Health Survey*, or on any questions of the year 8 *PE Survey*. There was a difference on just one of the 24 PE tasks, with students from intermediate schools scoring higher on *Ladder Ins and Outs* (p45).

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 was a difference on one of the 36 health tasks, with students from year 7 to 13 high schools scoring higher on *Link Task 8* (p30). There was also a difference on one of the 24 PE tasks, with students from intermediate schools scoring higher on *Link Task 11* (p46). There were no differences on any questions of the year 8 *Health Survey* or year 8 *PE Survey*.

### 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 34 health tasks: *Link Task 1* (p30) and *Link Task 23* (p55). On both of these tasks, students from small schools scored lowest. There were no differences on any of the 23 PE tasks, on any questions of the year 4 *Health Survey*, or on any questions of the year 4 *PE Survey*.

For year 8 students, there were differences on two of the 24 PE tasks, with students from large schools scoring highest (and students from medium-sized schools lowest) on *Racquet Strike* (p35) and *Ladder Ins and Outs* (p45). There were no differences on any of the 36 health tasks, and any questions of the year 8 *Health Survey*, or on any questions of the year 8 *PE Survey*.

### 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 one of the 34 health tasks and one of the 23 PE tasks. Students from rural areas scored lowest on *Clean Hands* (p28) and students from provincial towns lowest on *Leap* (p37). There were no differences on any questions of the year 4 *Health Survey* or the year 4 *PE Survey*.

For year 8 students, there was a difference among the three subgroups on one of the 24 PE tasks, with students from provincial towns scoring highest and students from main centres lowest on *Beanies* (p43). There were no differences on any of the 36 health tasks, on any questions of the year 8 *Health Survey*, or on any questions of the year 8 *PE Survey*.

### 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 two of the 34 health tasks. Students from regions of the North Island other than Auckland scored highest

on *Why Play?* (p20), while students from Auckland scored highest on *Link Task 23* (p55). There was also a difference on one of the 23 PE tasks, with students from the South Island scoring lowest on *Skipping Ropes* (p40). There were no differences on any questions of the year 4 *PE Survey*, but there was a difference on one question of the year 4 *Health Survey* (p63): students from the South Island indicated that their classes least often did things that helped them learn about health (question 7).

For year 8 students, there were differences among the three subgroups on five of the 36 health tasks, with students from the South Island highest on all five tasks: *Smoke Free* (p15), *Accidents* (p17), *School Lunches* (p18), *Agree or Disagree Y4* (p21), and *Role Models* (p53). There were also differences on four of the 24 PE tasks: students from the South Island scored lowest on *Racquet Strike* (p35) and *Poi Swings Y8* (p42), while students from Auckland scored lowest on *Beanies* (p43) and *Link Task 13* (p46). There were no differences on any questions of the year 8 *PE Survey*, but there was a difference on one question of the year 8 *Health Survey* (p63), with students from the South Island least positive about the value of learning about health (question 2).



### 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 14 of the 34 health tasks. Students in high decile schools scored higher than students in low decile schools on

all 14 tasks: *Accidents* (p17), *School Lunches* (p18), *Agree or Disagree Y4* (p21), *Infections* (p27), *Clean Hands* (p28), *Link Tasks 1, 2, 3, 4, 5 and 9* (p30), *Disappointment* (p52), *Playground Rules* (p58), and *Link Task 27* (p60). It is noteworthy that most of these tasks are in Chapter 3 (Personal Health). There were also differences on two questions of the year 4 *Health Survey* (p63), with students from low decile schools most positive about studying health at school (question 1) and reporting that their class more often did things that helped them learn about health (question 7).

There were differences on six of the 23 PE tasks: year 4 students from low decile schools scored highest on *Small Ball Catch* (p34), but lowest on *Foot Balance* (p38), *Bottom Balance* (p44), and *Link Tasks 12, 17 and 20* (p46). There was also a difference on one question of the year 4 *PE Survey* (p65), with students from medium decile schools thinking that their families were most positive about their capabilities in physical education.

For year 8 students, there were differences among the three subgroups on 16 of the 36 health tasks. Students in high decile schools performed better than students in low decile schools on all 16 tasks: *Smoke Free* (p15), *Being Healthy* (p16), *School Lunches* (p18), *Agree or Disagree Y8* (p23), *Infections* (p27), *Listen to Your Heart!* (p29), *Link Tasks 1, 2, 3, 4, 5, 6 and 9* (p30), *Good Neighbours* (p57), *Fair Play* (p59), and *Link Task 26* (p60). It is noteworthy that none of these tasks are in Chapter 5 (Relationships with Other People). There were differences on two questions of the year 8 *Health Survey* (p63), with students from low decile schools most positive about the value of learning about health (question 2) and about learning more health as they got older (question 6).

There were differences on eight of the 24 PE tasks, with year 8 students from low decile schools scoring lower than students from high decile schools on all 8 tasks: *Small Ball Catch* (p34), *Racquet Strike* (p35), *Leap* (p37), *Foot Balance* (p38), and *Link Tasks 12, 13, 17 and 20* (p46). There were no differences on any questions of the year 8 *PE Survey*.

## 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 29 health tasks was 0.09 (girls averaged 0.09 standard

deviations higher than boys). This indicates a small difference, on average. The mean effect size was very small (0.04) for Chapter 3 tasks, but larger (0.16) for tasks in Chapters 5 and 6. There were differences on five of the 29 tasks: boys scored higher on *Link Task 1* (p30), but girls scored higher on *What Do You Think?* (p48), *Jamie* (p54), *Link Task 22* (p55) and *Good Neighbours* (p57). There were no differences on any question of the year 4 *Health Survey*.

The mean effect size across the 22 PE tasks was 0.10 (year 4 boys averaged 0.10 standard deviations higher than girls). This indicates a small difference, on average. There were statistically significant differences on 15 of the 22 tasks. Boys scored higher on nine tasks: *Run* (p32), *Dodge* (p33), *Small Ball Catch* (p34), *Racquet Strike* (p35), *Distance Throw* (p36), *Leap* (p37), and *Link Tasks 10, 11 and 19* (p46). Girls scored higher on six tasks: *Foot Balance* (p38), *Skipping Ropes* (p40), *Poi Swings Y4* (p41), *Bottom Balance* (p44), *Ladder Ins and Outs* (p45) and *Link Task 17* (p46). There was also a difference on one question of the year 4 *PE Survey* (p65): boys reported a greater amount of physical exercise over the 24 hours before completing the survey (question 9).

For year 8 students, the mean effect size across the 32 health tasks was 0.20 (girls averaged 0.20 standard deviations higher than boys): a moderate difference. There were statistically significant differences favouring girls on 13 of the 32 tasks: *Smoke Free* (p15), *Why Play?* (p20), *Link Tasks 4, 6 and 9* (p30), *What Do You Think?* (p48), *Suzy* (p50), *Link Tasks 22 and 23* (p55), *Good Neighbours* (p57), *Playground Rules* (p58), *Fair Play* (p59) and *Link Task 27* (60). There were also differences on two questions of the year 8 *Health Survey* (p63). Girls thought that they were better at health (question 4) and were more positive about learning more about health as they got older (question 3).

The mean effect size across the 23 PE tasks was 0.10 (year 8 boys averaged 0.10 standard deviations higher than girls). This indicates a small difference, on average. There were statistically



significant differences on 11 of the 23 tasks. Boys scored higher on seven tasks: *Run* (p32), *Small Ball Catch* (p34), *Racquet Strike* (p35), *Distance Throw* (p36), and *Link Tasks 10, 11 and 16* (p46). Girls scored higher on four tasks: *Skipping Ropes* (p40), *Poi Swings Y8* (p42), *Ladder Ins and Outs* (p45) and *Link Task 17* (p46). There were also difference on three questions of the year 8 *PE Survey* (p65): boys were more positive about doing PE at school (question 1), how good they thought they were at PE (question 2) and wanting to do more PE (question 7).

### 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 29 health tasks was 0.25 (Pakeha students averaged 0.25 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences ( $p < .01$ ) on nine of the 29 tasks, with Pakeha students higher on all nine tasks: *Link Tasks 1, 2, 3, 4, 5 and 9* (p30), *Link Task 23* (p55), *Good Neighbours* (p57) and *Link Task 26* (p60). There was a difference on one question of the year 4 *Health Survey* (p63): Māori students reported that their class more often did things to help them learn about health (question 7).

The mean effect size across the 22 PE tasks was 0.09 (year 4 Māori students averaged 0.09 standard deviations higher than Pakeha students). This is a small difference. There were statistically significant differences, all favouring Māori students, on four of the 22 tasks: *Small Ball Catch* (p34) *Hoops* (p39), *Skipping Ropes* (p40) and *Poi Swings Y4* (p41). There were no differences on any questions of the year 4 PE Survey.

For year 8 students, the mean effect size across the 32 health tasks was 0.23 (Pakeha students averaged 0.23 standard deviations higher than Māori students). This is a moderate difference. There were statistically significant differences ( $p < .01$ ) on nine of the 32 tasks, with Pakeha students higher on all nine tasks: *Being Healthy* (p16), *Accidents* (p17), *Listen to Your Heart!* (p29), *Link Tasks 2, 4, 6 and 8* (p30), *Link Task 21* (p55) and *Link Task 26* (p60). There were no differences on questions of the year 8 Health Survey.

The mean effect size across the 23 PE tasks was 0.06 (year 8 Māori students averaged 0.06 standard deviations higher than Pakeha students). This is a small difference. There were statistically significant differences on six of the 23 tasks. Māori students scored higher on four tasks: *Skipping Ropes* (p40), *Poi Swings Y4* (p41) and *Link Tasks 10 and 15* (p46). Pakeha students scored higher on two tasks: *Foot Balance* (p38) and *Link Task 12* (p46). There were also differences on two questions of the year 8 PE Survey (p65). Māori students were more enthusiastic about doing additional PE (question 7) and about continuing to learn PE as they got older (question 8).

### 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 29 health tasks was 0.26

(Pakeha students averaged 0.26 standard deviations higher than Pasifika students). This is a moderate difference. The difference was larger for personal health tasks (Chapter 3), where the mean effect size was 0.35, and smaller for the tasks of Chapters 5 and 6, where the mean effect size was 0.13. There were statistically significant differences on 10 of the 29 tasks, with Pakeha students higher on all 10 tasks: *Smoke Free* (p15), *Accidents* (p17), *School Lunches* (p18), *Clean Hands* (p28), *Link Tasks 1, 4, 5, 6, and 9* (p30), and *Link Task 26* (p55). All except the last task were in Chapter 3 (Personal Health). There were also differences on four questions of the year 4 Health Survey (p63): Pasifika students were more positive about doing health at school (question 1), learning more about health as they got older (question 3), and reported that their class more often did things that helped them learn about health (question 7), but Pakeha students thought that learning about health was more useful to them (question 2).

The mean effect size across the 22 PE tasks was 0.09 (year 4 Pasifika students averaged 0.09 standard deviations higher than Pakeha students). This is a small difference. There were statistically significant differences on 10 of the 22 tasks. Pasifika students scored higher on seven tasks: *Small Ball Catch* (p34), *Hoops* (p39), *Skipping Ropes* (p40), and *Link Tasks 15, 16, 18 and 19* (p46). Pakeha students scored higher on three tasks: *Foot Balance* (p38), *Bottom Balance* (p44) and *Link Task 20* (p46). There were also differences



on two questions of the year 4 PE Survey (p65): Pasifika students were more positive about doing PE at school (question 1) and about doing additional PE (question 7).

For year 8 students, the mean effect size across the 32 health tasks was 0.32 (Pakeha students averaged 0.32 standard deviations higher than Pasifika students). This is a moderate difference. The difference was larger for personal health tasks (Chapter 3), where the mean effect size was 0.41, and smaller for the tasks of Chapters 5 and 6, where the mean effect size was 0.19. There were statistically significant differences ( $p < .01$ ) on 19 of the 32 tasks, with Pakeha students higher on all 19 tasks: fifteen of the 19 tasks in Chapter 3, plus *Suzu* (p50), *Good Neighbours* (p57) and *Link Tasks 26 and 27* (p60). There were no differences on questions of the year 8 Health Survey.

The mean effect size across the 23 PE tasks was 0.10 (year 8 Pakeha students averaged 0.10 standard deviations higher than Pasifika students). This is a small difference. There were statistically significant differences on six of the 23 tasks. Pasifika students scored higher on *Small Ball Catch* (p34), while Pakeha students scored higher on five tasks: *Leap* (p37), *Beanies* (p43), and *Link Tasks 12, 13, and 20* (p46). There were also differences on two questions of the year 8 PE Survey (p65). Pasifika students thought that they were better at PE (question 2) and were more positive about trying things in PE that they hadn't done before (question 5).

### 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 29 health tasks was 0.08 (students for whom English was the predominant language at home averaged 0.08 standard deviations

higher than the other students). This is a small difference. There were statistically significant differences on four of the 29 tasks. Students for whom English was the predominant language at home scored higher on *Smoke Free* (p15), *Accidents* (p17), *Clean Hands* (p28) and *Link Task 8* (p30). There were also differences on three questions of the year 4 *Health Survey* (p63). Students for whom the predominant language at home was not English were more positive about doing health at school (question 1) and learning more about health as they got older (question 3), and thought that their class more often did things that helped them learn about health (question 7).

The mean effect size across the 22 PE tasks was 0.08 (year 4 students for whom English was the predominant language at home averaged 0.08

standard deviations higher than the other students). This is a small difference. There were statistically significant differences on two of the 22 tasks. Students for whom English was the predominant language at home scored higher on *Ladder Ins and Outs* (p45) and *Link Task 18* (p46). There was also a difference on one question of the year 4 *PE Survey* (p65). Students for whom the predominant language at home was English reported doing a greater amount of vigorous physical exercise in the 24 hours before the survey (question 9).

For year 8 students, the mean effect size across the 32 health tasks was 0.20 (students for whom English was the predominant language at home averaged 0.20 standard deviations higher than the other students). This is a moderate difference. There were

statistically significant differences on five of the 32 tasks. Students for whom English was the predominant language at home scored higher on *Accidents* (p17), *School Lunches* (p18), *Listen to Your Heart!* (p29), *Link Task 22* (p55) and *Link Task 27* (p60). There were no differences on any questions of the year 8 *Health Survey*.

The mean effect size across the 23 PE tasks was 0.03 (year 8 students for whom English was the predominant language at home averaged 0.03 standard deviations higher than the other students). This is a negligible difference. There was a statistically significant difference on one of the 23 tasks: students for whom English was the predominant language at home scored higher on *Leap* (p37). There were no differences on any question of the year 8 *PE Survey*.



### Summary, with Comparisons to Previous Health and Physical Education 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 health or PE tasks at either year level. The same was true for the 2002 and 1998 assessments.

There were statistically significant differences in the performance of students from low, medium and high decile schools on 41 percent of the health tasks at year 4 level (compared to 32 percent in 2002 and 44 percent in 1998), and 44 percent of the health tasks at year 8 level (compared to 44 percent in 2002 and 38 percent in 1998). For the PE tasks, there were differences on 26 percent of the tasks at year 4 level (compared to five percent in 2002 and 17 percent in 1998), and 33 percent of the tasks at year 8 level (compared to eight percent in 2002 and 17 percent in 1998).

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 slightly higher than boys on health tasks, with a mean effect size of 0.09 (exactly the same as in 2002). Year 8 girls averaged moderately higher than boys on health tasks, with a mean effect size of 0.20 (little different from 0.17 in 2002). On the PE tasks, year 4 boys averaged a little higher than girls, with a mean effect size of 0.10 (slightly reduced from 0.15 in 2002). Year 8 boys also averaged slightly higher than girls on PE tasks, with a mean effect size of 0.10 (exactly the same as in 2002). Boys did better on tasks that involved physical strength or kicking, hitting, catching or throwing balls, while girls did better on some of the other tasks (such as skipping, poi, balancing and patterned movement).

Pakeha students averaged moderately higher than Māori students on the health tasks, with mean effect sizes of 0.25 for year 4 students (slightly increased from 0.20 in 2002) and 0.23 for year 8 students (exactly the same as in 2002). On the PE tasks, however, Māori students scored slightly higher than Pakeha students at both year

levels. The mean effect size for year 4 students was 0.09 (slightly reduced from 0.14 in 2002), while for year 8 students the mean effect size was 0.06 (also slightly reduced from 0.10 in 2002).

Pakeha students averaged moderately higher than Pasifika students on the health tasks, with mean effect sizes of 0.26 for year 4 students and 0.32 for year 8 students (revealing substantially reduced disparities of performance compared to 2002, when the two effect sizes were 0.40 and 0.45). On the PE tasks, Pasifika students averaged a little higher than Pakeha students at year 4 level (mean effect size of 0.09, reduced from 0.17 in 2002), but the converse was true at year 8 level (mean effect size of 0.10 favouring Pakeha students, increased from 0.00 in 2002).

Compared to students for whom the predominant language at home was not English, students from homes where English predominated averaged slightly higher at year 4 level (mean effect size 0.08 for both health and physical education tasks) and on year 8 level physical education tasks (mean effect size of 0.03) Their advantage was greater on year 8 health tasks (mean effect size of 0.20). Comparative figures are not available for the assessments in 2002.