

# CHAPTER 4: New Zealand Year 9 Students' Science Achievement

## KEY POINTS

- The mean science score achieved by New Zealand Year 9 students in 1998 was almost identical to the mean science score achieved by that of their Year 9 counterparts in 1994.
- The statistically significant difference between the mean science scores of New Zealand Year 9 boys and girls observed in 1994 (in favour of boys) no longer existed in 1998.
- There was no statistically significant difference in the mean science scores of New Zealand Year 9 boys in 1994 and 1998; the same result was true for girls.
- There was no difference in mean science scores achieved by any ethnic grouping of New Zealand Year 9 students between 1994 and 1998. The mean scores of the Pakeha/European and Asian ethnic groupings continued to be significantly above the mean scores for the Maori and Pacific ethnic groupings.
- *Scientific Inquiry and the Nature of Science* was a content area of relative strength for New Zealand Year 9 students.
- New Zealand Year 9 girls achieved at least nominally above the international means for girls in all six science content areas that were examined in 1998. In the areas of *Physics* and *Scientific Inquiry and the Nature of Science* these higher scores were statistically significant.

## INTRODUCTION

This chapter presents the science achievement results for New Zealand Year 9 students in TIMSS-98 in a national context. In addition to examining the results by gender and by ethnic grouping, whenever possible, comparisons are made with the TIMSS-94 Year 9 student achievement data. In the latter half of the chapter, Year 9 students' achievements in each of the six separate science content areas are examined in both a national and international context.

## OVERALL ACHIEVEMENT RESULTS

Firstly, to recap: TIMSS-98 was designed to measure trends in mathematics and science achievement at the lower secondary level. It is also, in part, a quasi-longitudinal study of a cohort: the cohort assessed in TIMSS-94 as Year 5 students was again assessed in TIMSS-98 as Year 9 students (although they were not necessarily the same individuals in each study). While it is not possible to *directly* compare the science achievement of the cohort as Year 5 students in TIMSS-94 to the cohort's achievement in science as Year 9 students in TIMSS-98 (because of the independent assessments and scales involved) there are a number of methods that facilitate comparative analysis. For example, relativities within the cohort that were observed in the 1994 and 1998 can be compared. Further to this, many aspects of the science achievement of Year 9 students in 1998 can be compared with those of their Year 9 counterparts in TIMSS-94.

### The overall trend from 1994 to 1998

The most notable feature of New Zealand's science achievement between TIMSS-94 and TIMSS-98 is the distinct lack of change — both in the overall mean score and the overall distribution.

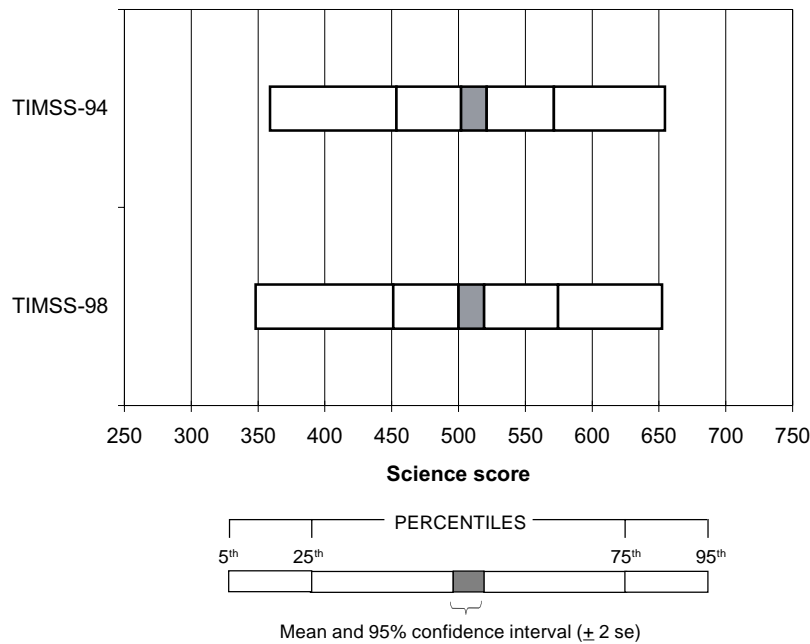
In TIMSS-94 and again in TIMSS-98, the mean science score attained by New Zealand students was about 510.

**TABLE 4.1: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998**

Year 9 students	Mean science scale score (se)	
	1994	1998
Overall mean	511 (4.9)	510 (4.9)

*(se) Standard errors appear in parentheses. (Also refer to TN.3 in Technical Notes.) We can say with 95% confidence that the true population mean lies within two standard errors of the observed (reported) mean.*

Figure 4.1 presents the distribution of science scores for Year 9 students in 1994 and 1998 (see also Table C.1, Appendix C). For an explanation on the interpretation of the percentiles, readers can refer to page 37 of this report.

**FIGURE 4.1: THE DISTRIBUTION OF YEAR 9 STUDENTS' SCIENCE SCORES IN 1994 AND 1998**

Note: see boxed text on page 37 for a guide to interpreting percentiles.

Note that amongst countries participating in both TIMSS-94 and TIMSS-98 there was no uniform pattern in trends in science achievement. For example, unlike in New Zealand, in five countries there were statistically significant differences in mean science achievement between the two studies (of these, four observed an increase in students' science achievement in 1998, the other a decrease; see Chapter 2 and Martin et al, 2000 for more detail).

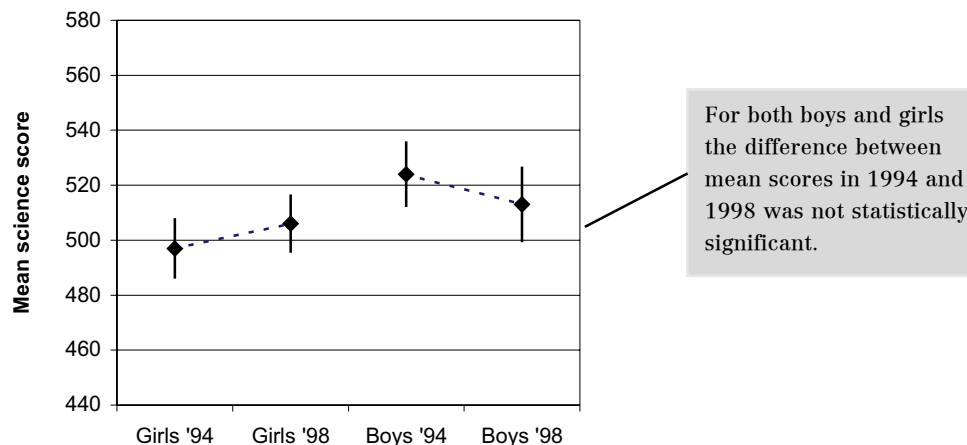
As noted above, the New Zealand cohort assessed in TIMSS-98 as Year 9 students was also assessed in TIMSS-94 as Year 5 students. Because similar cohort assessment occurred in several other countries that participated in both TIMSS-94 and TIMSS-98 (referred to as 'trends' countries) it was possible to compare the progress of the New Zealand cohort with cohorts in other countries over the four year period. The trend in relative science achievement in New Zealand of this cohort from 1994 to 1998 is discussed in Chapter 2. To briefly reiterate, it was found that in both TIMSS-94 and TIMSS-98 the mean science score achieved by the New Zealand cohort did not differ significantly from the mean score for the 17 'trend' countries. In other words, there was little change in this New Zealand cohort's mean science achievement relative to its international counterparts as they progressed from the middle primary to the lower secondary school level. Martin et al (2000) and Chapter 2 of this report provide descriptions of relative international comparisons across the two studies.

## GENDER AND SCIENCE ACHIEVEMENT

In TIMSS-98 the mean science score of Year 9 boys did not differ significantly<sup>1</sup> from that of girls. This contrasts with TIMSS-94 where the mean score of Year 9 boys was significantly higher than that for girls<sup>2</sup>. Figure 4.2 plots the mean scores of Year 9 boys and girls in both TIMSS-94 and TIMSS-98 (see Table C.2 in Appendix C for detail of the means and standard errors).

<sup>1</sup> Throughout this chapter, differences are described as 'significant', or 'statistically significant' where significance tests meet  $\alpha = 0.05$  (ie, the 5% level); see also TN.4 in Technical Notes.

<sup>2</sup> There was a small effect size of 0.3 between New Zealand boys and girls science achievement in TIMSS-94. See TN.6 in Technical Notes for an outline of effect sizes.

**FIGURE 4.2: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY GENDER**

*The data points are the mean scores. The vertical lines extending from the data point show the 95% confidence interval around the mean, ie,  $\pm 2$  standard errors.*

Interestingly, however, the mean score for Year 9 boys in TIMSS-98 was not statistically significantly different from the mean score achieved by Year 9 boys in TIMSS-94. Similarly, there was no statistically significant difference between the mean science scores achieved by Year 9 girls in each of TIMSS-94 and TIMSS-98.

When the cohort assessed in TIMSS-98 as Year 9 students was assessed in TIMSS-94 as Year 5 students, there was no statistical significance in the difference between the mean science scores achieved by Year 5 boys and girls (although there was a nominal difference in favour of girls) (Martin et al, 1997; Chamberlain 1997a). It is interesting to note, too, that the lack of significant gender difference for the Year 5 student cohort has been carried through into 1998 when the students were now Year 9 (although, again, there was a nominal difference — this time in favour of boys).

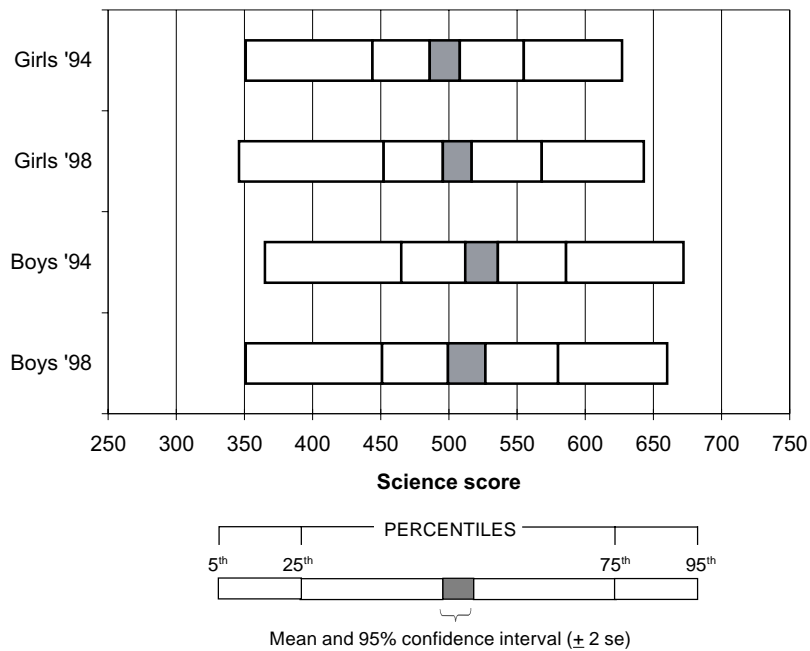
Despite appearances to the contrary, this lack of overall gender difference in TIMSS-98 is not at odds with gender differences in science performance found at the Year 8 level in the 1999 National Education Monitoring Project (NEMP) Science study. In NEMP, Crooks and Flockton (2000) reported gender differences favouring boys on 13 of 48 science items. In TIMSS-98, there were also gender differences at the item level but they are not the subject of this report. Rather, the focus here is on overall science achievement. The caveat is that, when looking at the results of the NEMP and the current TIMSS-98 results in tandem, one must consider the different methodologies and reporting procedures: whereas NEMP analysed and reported differences on discrete assessment tasks, TIMSS-98 initially used Item Response Theory to integrate a range of assessment and contextual information from students into an overall science scale which estimates latent traits<sup>3</sup> of population groups: in this case 'general science ability'. This does not, of course, preclude item level analyses of TIMSS-98 in the future.

There are a few notable differences in the distributions of the Year 9 science scores by gender for each of TIMSS-94 and TIMSS-98 (Figure 4.3). For girls, on the whole, the distribution translated positively from 1994 to 1998. However, concomitant to this, is the greater negative skew in the girls' distribution for 1998. For boys, the distribution has translated negatively from 1994 to 1998 and this translation is

<sup>3</sup> A latent trait is one that is underlying and not directly observable.

fairly uniform. In TIMSS-98 there were only small (non-significant) differences between the proportions of girls and boys reaching the overall national median (49% for girls, 51% for boys) and the overall national upper quartile (23% for girls, 27% for boys) (Martin et al, 2000). See Table C.3 in Appendix C for the details of the percentiles.

**FIGURE 4.3: THE DISTRIBUTION OF YEAR 9 STUDENTS' SCIENCE SCORES IN 1994 AND 1998, BY GENDER**



## ETHNICITY AND SCIENCE ACHIEVEMENT

No Year 9 ethnic grouping achieved a mean science score in 1998 that was statistically significantly different from its Year 9 mean score in 1994 (see Table 4.2). Note that due to the small overall number and wide diversity within the 'other' ethnic grouping, this group will not be considered in further analyses in this chapter.

**TABLE 4.2: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY ETHNIC GROUPING**

Year	Mean science scale score (se)				
	Pakeha/ European	Maori	Pacific	Asian	Other ethnic groups
1994	533 (4.4)	472 (5.6)	430 (8.5)	498 (12.0)	526 (12.0)
1998	534 (4.5)	472 (6.0)	430 (12.0)	515 (9.9)	513 (17.3)

*Note: see Chapter 5 for proportions of students in each ethnic grouping.*

*(se) Standard errors appear in parentheses.*

A corollary of the lack of change in mean science scores within the ethnic groupings is that the relativity of achievement between Year 9 ethnic groupings observed in 1994 was maintained in 1998. Table 4.3 shows where the differences were statistically significant (note that the table also applied for 1994).

#### Reading the significance table

Table 4.3 is read so that an arrow pointing upwards indicates the group in the **row** has a significantly higher mean score than the corresponding group in the **column** (eg, Pakeha/European had a significantly higher mean than Maori). A downward arrow indicates a significantly lower mean score (of the row group relative to the column group), and a circle indicates no significant difference between the groupings.

**TABLE 4.3: COMPARISON OF YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY ETHNIC GROUPING**

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		▲	▲	●
Maori	▼		▲	▼
Pacific	▼	▼		▼
Asian	●	▲	▲	

▲	Mean achievement was statistically significantly higher
●	No difference in mean achievement
▼	Mean achievement was statistically significantly lower

*Note:* Significance tests adjusted for multiple comparisons. See TN.5 in Technical Notes. Also refer to Table C.4 in Appendix C for *t* statistic of differences.

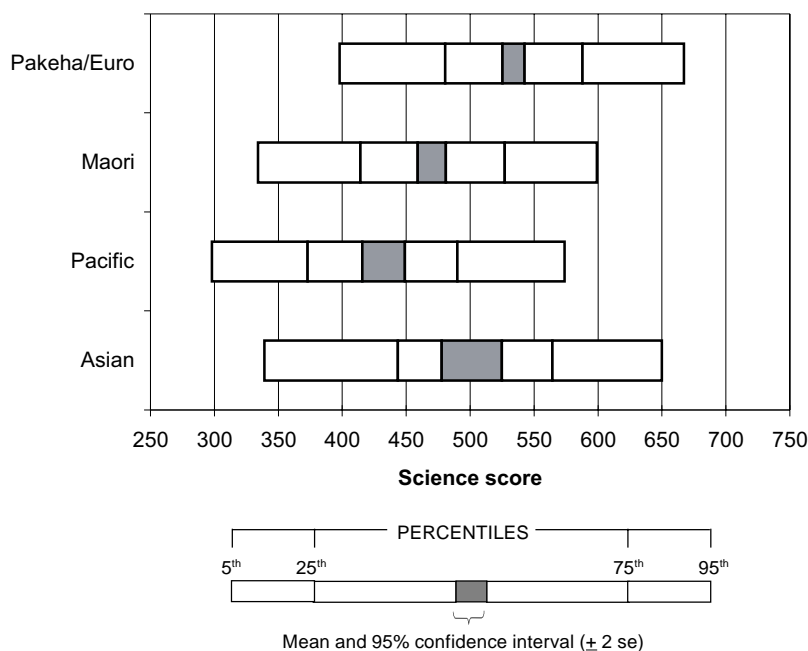
Notwithstanding the lack of movement in relative differences of the mean scores achieved by students in the four main ethnic groupings between the two studies there was some shift in the magnitude of differences in relative science achievement — but only when groups were compared with the Asian ethnic grouping. For example, the effect size between Pakeha/European and Asian has reduced from 0.44 in 1994 to 0.22 in 1998. Similarly the effect size between Asian and Maori has increased from 0.31 to 0.47 (see TN.6 in Technical Notes on effect sizes). These figures indicate relatively stronger science achievement among the Asian Year 9 student population in 1998 compared with 1994.

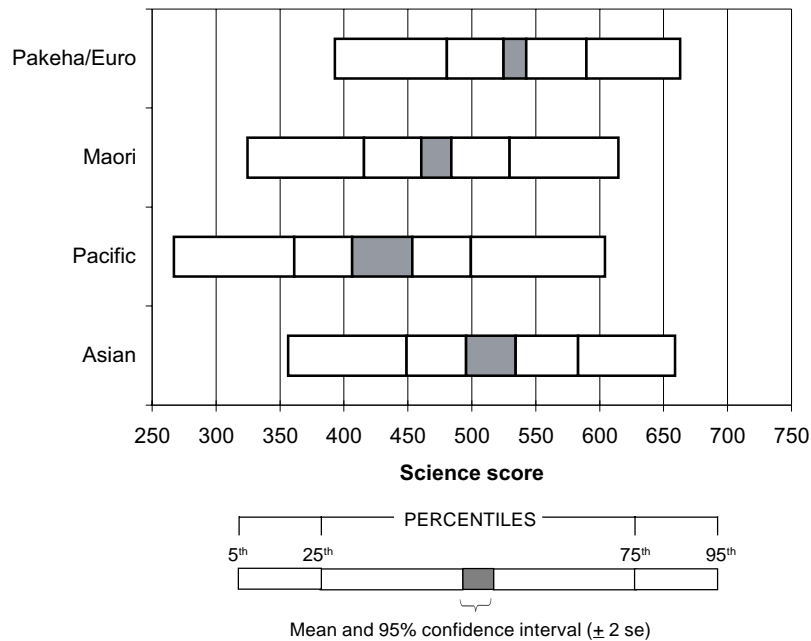
Another way to examine trends in achievement by ethnicity is to look at relative performance at the Year 5 level in 1994 and compare it with relative performance at the Year 9 level in 1998. A good way to do this is to again compare the effect sizes between ethnic grouping pairings. The examination of effect sizes revealed that while the relative science achievement of Maori and Pakeha/European remained stable from Year 5 in 1994 through to Year 9 in 1998, Asian students' performance improved relative to all other ethnic groupings and Pacific students' achievement declined. One thing to consider about these findings, however, is that the Asian and Pacific populations are typically subject to greater migration effects than Pakeha/European or Maori — in other words, there could be greater variability in the make-up of these cohorts over time which could impact on achievement (see Tables C.5a and C.5c in Appendix C for the effect sizes between the four main Year 5 and Year 9 ethnic groupings in 1994 and 1998 respectively).

Figures 4.4a and 4.4b show the distributions of Year 9 students' science scores by the four main ethnic groupings in 1994 and 1998 respectively. In general, there was little change in the distributions between the two studies. The Pacific and Asian populations exhibited the most change. For Pacific, although the mean score remained about the same, there was a tendency in 1998 for scores to show greater variability around the mean than they had in 1994: the stronger students scored at a slightly higher level in 1998 and the weaker students scored slightly lower. For Asian students, there was small but general positive shift in the distribution of science scores, with the stronger, average, and weaker students all performing slightly better than their 1994 counterparts. However, it is important to note there are greater errors around the percentile estimates for Pacific and Asian populations (because of their relatively small samples) which must be considered when interpreting these results.

Among other things, these figures also illustrate the heterogeneity of the sub-populations, with both high-scoring and low-scoring individuals present in every ethnic grouping. This point, often obscured when focusing on differences in mean scores, should always be borne in mind when examining achievement differences between groups. For example, while the mean science achievement score for Pacific students in 1998 was 430, the five percent of best performing Pacific students achieved a score of over 600, whereas the five percent of lowest performing students scored no more than 270 (see Table C.6 in Appendix C for details of the percentiles for Year 9 students in each ethnic grouping in 1994 and 1998).

**FIGURE 4.4A: THE DISTRIBUTION OF YEAR 9 STUDENTS' SCIENCE SCORES IN 1994, BY ETHNIC GROUPING**



**FIGURE 4.4B: THE DISTRIBUTION OF YEAR 9 STUDENTS' SCIENCE SCORES IN 1998, BY ETHNIC GROUPING**

### Ethnicity, gender and science achievement

In a recent extensive literature review of gender differences in the New Zealand school sector, Alton-Lee and Praat (2000, p87) summarised that:

*Gender, ethnicity and social class intersect in complex ways to influence the ways students negotiate curriculum ...*

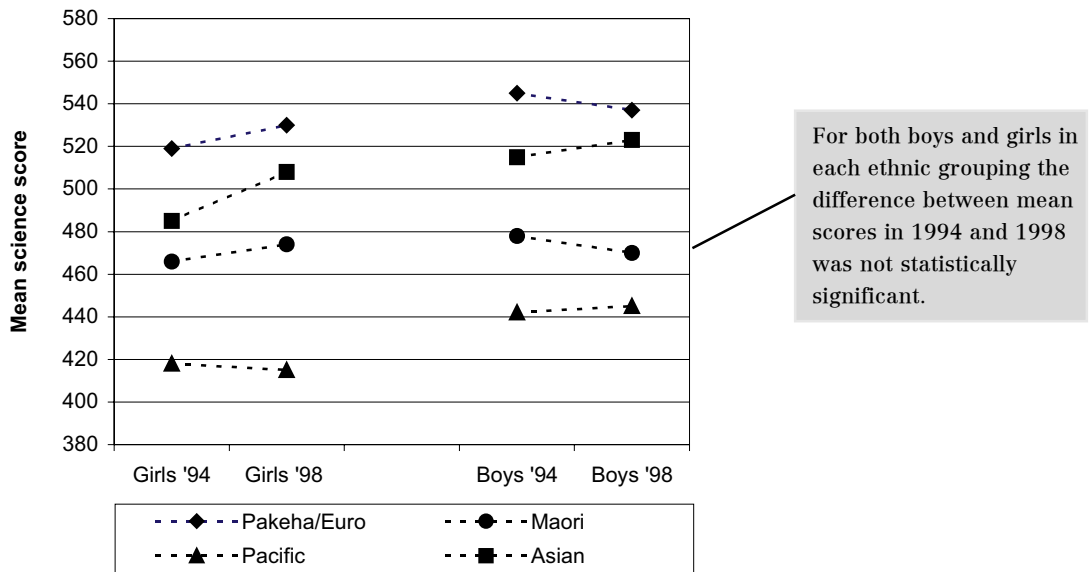
Ignoring, for the present, the social class factor, the Year 9 student science achievement results in both 1994 and 1998 reinforce the above proposition. Table 4.4 presents the 1998 mean science scores for Year 9 boys and girls by ethnic grouping (Table C.7 Appendix C provides equivalent information for 1994). Note that within each ethnic group there was no statistically significant difference in mean science scores achieved by girls and boys.

**TABLE 4.4: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY ETHNIC GROUPING AND GENDER**

1998 assessment	Mean science scale scores (se)							
	Pakeha/European		Maori		Pacific		Asian	
	girls	boys	girls	boys	girls	boys	girls	boys
	530 (4.8)	537 (6.9)	474 (6.8)	470 (7.1)	415 (15.3)	445 (15.4)	508 (13.2)	523 (13.9)
Overall mean 534 (4.5)		472 (6.0)		430 (12.0)		515 (9.9)		

(se) Standard errors appear in parentheses.

Figure 4.5 shows the mean science scores of students grouped by ethnicity and gender in both studies. While, for some 'gender-ethnic sub-populations', the differences in scores between the two studies appear considerable (eg, for Asian girls), none were found to be statistically significant.

**FIGURE 4.5: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY ETHNIC GROUPING AND GENDER**

### SOCIO-ECONOMIC FACTORS AND SCIENCE ACHIEVEMENT

It is well documented that low socio-economic status (of both families and communities) is associated with relatively poor science performance in school (eg, Beaton et al 1996a; Chamberlain 1996a; Walker 1998; Crooks & Flockton, 2000).

One useful way of characterising the socio-economic features of the school community is to use the Ministry of Education's indicator Targeted Funding for Educational Achievement (TFEA)<sup>4</sup>. The mean science scores for Year 9 students attending schools in each of three aggregated decile bands are shown in Table 4.5. (Table C.8, Appendix C provides the information for 1994.)

**TABLE 4.5: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY SCHOOLS' TARGETED FUNDING FOR EDUCATIONAL ACHIEVEMENT (TFEA) DECILE BAND**

Schools' TFEA decile band	1998	
	% of students	Mean science score (se)
Low (deciles 1-3)	18	471 (11.0)
Medium (deciles 4-7)	48	498 (6.6)
High (deciles 8-10)	30	544 (7.5)
No TFEA indicator (ie, private schools)	4	558 (8.7)

(se) Standard errors appear in parentheses.

See footnote 1 of this chapter on TFEA deciles.

<sup>4</sup> The TFEA indicator is not entirely a measure of the socio-economic features of the school's community and some anomalies may exist. However, it is robust enough to be indicative of the socio-economic milieu of schools' intakes and any anomalies are ameliorated in the current analysis by grouping the deciles into low, medium and high.

Those schools with no decile were exclusively independent (private) schools in the 1998 study. However, it is worth noting that the high performance of students in these schools is probably confounded by their relatively high socio-economic status of their family/community backgrounds. The mean score of students in 'no decile' schools was not significantly higher than that of the students in high decile schools.

#### Reading the significance table

Table 4.6 is read so that an arrow pointing upwards indicates the group in the row has a significantly higher mean score than the corresponding group in the column (eg, students in high TFEA decile band schools had a significantly higher mean than students in low TFEA decile band schools). A downward arrow indicates a significantly lower mean score (of the row group relative to the column group), and a circle indicates no significant difference between the groupings.

**TABLE 4.6 COMPARISON OF YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY SCHOOLS' TARGETED FUNDING FOR EDUCATIONAL ACHIEVEMENT (TFEA) DECILE BAND**

Schools' TFEA decile band	Low (deciles 1-3)	Medium (deciles 4-7)	High (deciles 8-10)	No TFEA indicator
Low (deciles 1-3)		●	▼	▼
Medium (deciles 4-7)	●		▼	▼
High (deciles 8-10)	▲	▲		●
No TFEA indicator	▲	▲	●	

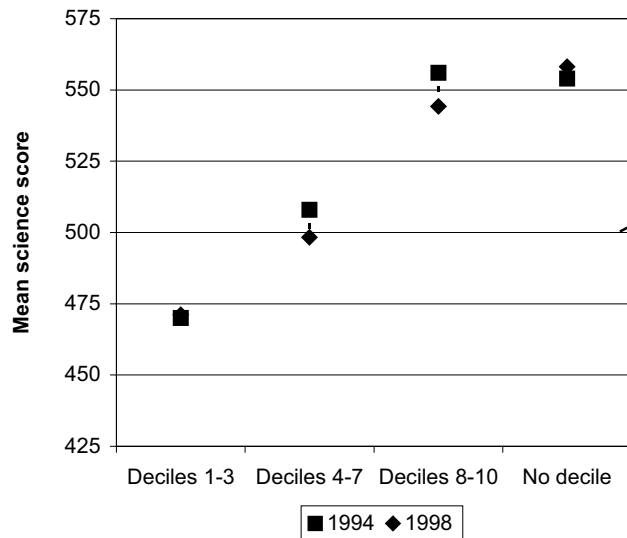
▲	Mean achievement was statistically significantly higher
●	No difference in mean achievement
▼	Mean achievement was statistically significantly lower

*Note: Significance tests adjusted for multiple comparison of means. See TN.5 in Technical Notes. Also refer to Table C.10 in Appendix C for t statistic for comparing differences.*

*Refer to page 44 for an outline of the TFEA decile bands.*

It is also important to remember that although there are associations between the socio-economic status of the community and the science achievement of students, there are high and low achieving students throughout all types of schools and communities. For example, five percent of students in the high decile band schools scored no more than 400 on the science achievement scale. Conversely, five percent of students in the low decile band schools scored at least 635. For details on the distribution of science scores by schools' TFEA decile band see Table C.9, Appendix C)

As Figure 4.6 shows, there was a similar pattern of relativity in mean scores in TIMSS-94. Within each TFEA decile band, there was no statistically significant difference found in the mean science scores achieved by students between 1994 and 1998.

**FIGURE 4.6: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY SCHOOLS' TARGETED FUNDING FOR EDUCATIONAL ACHIEVEMENT (TFEA) DECILE BAND**

For students attending schools in each of the TFEA decile bands the difference between mean scores in 1994 and 1998 was not statistically significant.

Refer to page 44 for an outline of the TFEA decile bands.

## SCHOOL TYPE AND SCIENCE ACHIEVEMENT

When science achievement was examined on the basis of school type — coeducational, single-sex boys', or single-sex girls' schools — an interesting picture emerged (Table 4.7). At an overall level there was only one statistically significant difference in mean science scores for students in the three school types and that was the difference in achievement between students in single-sex girls' schools and those in coeducational schools.

**TABLE 4.7 YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY SCHOOL TYPE**

School type	1998	
	% of students	Mean science score (se)
Coeducational	68	499 (5.5)
Single-sex boys'	16	530 (17.6)
Single-sex girls'	17	532 (11.0)

(se) Standard errors appear in parentheses.

However, when the coeducational students were grouped by gender, the analysis revealed that both boys and girls who attended single-sex schools achieved significantly higher mean science scores than girls attending coeducational schools. Table 4.8 shows the statistical significance of differences (adjusted for multiple comparisons) between the 1998 mean science score by gender and school type.

**Reading the significance table**

Table 4.8 is read so that an arrow pointing upwards indicates the group in the **row** has a significantly higher mean score than the corresponding group in the **column** (eg, girls in single-sex schools had a significantly higher mean than girls in co-educational schools). A downward arrow indicates a significantly lower mean score (of the row group relative to the column group), and a circle indicates no significant difference between the groupings.

**TABLE 4.8: COMPARISON OF YEAR 9 STUDENTS' MEAN SCIENCE ACHIEVEMENT SCORES IN 1998, BY GENDER AND SCHOOL TYPE**

Gender	School type	Girls		Boys	
		Coed	Single-sex	Coed	Single-sex
Girls	Coed		▼	●	▼
	Single-sex	▲		●	●
Boys	Coed	●	●		●
	Single-sex	▲	●	●	

▲	Mean achievement was statistically significantly higher
●	No difference in mean achievement
▼	Mean achievement was statistically significantly lower

*Note: Significance tests adjusted for multiple comparison of means. See TN.5 in Technical Notes. Refer to Table C.11 in Appendix C for *t* statistic for comparing differences.*

It is very important to note that, just as for mathematics, it is likely that the socio-economic characteristics of a school's intake is a confounding factor in the differences in science achievement across school type. This can be simply illustrated by a cursory examination of science achievement alongside two proxy socio-economic indicators — firstly, the Ministry of Education's Targeted Funding for Educational Achievement (TFEA decile) school level indicator; and, secondly, a student level proxy for wealth, the Home Education Resources (HER) Index (this index is described in Chapter 5).

If the science achievement of students is examined alongside school type and TFEA decile, the type of school has less association with achievement than the decile of the school. For example, when the science achievement of the Year 9 students within schools grouped by type and TFEA decile band is compared, no statistically significant differences between the mean science scores of students in different school types within either the medium or high decile bands were found<sup>5</sup>. Yet, students in high decile band schools consistently achieve statistically significantly higher mean science scores than those in medium decile band schools of the same school type.

In our second illustration, science achievement is considered alongside school type and a proxy student-level measure of wealth: namely, the Home Education Resources (HER) Index developed for TIMSS-98. As described more fully in Chapter 5, the HER Index has three levels: high, medium and low, whereby categorisation as high on the index corresponds to a relatively higher level of educational resources within the student's home. When the science achievement of these medium level HER<sup>6</sup> students was

<sup>5</sup> Around half of students in each of the three school types were in medium decile (4-7) band and between a quarter and a half of students in each of the three school types were in high decile (8-10) band schools.

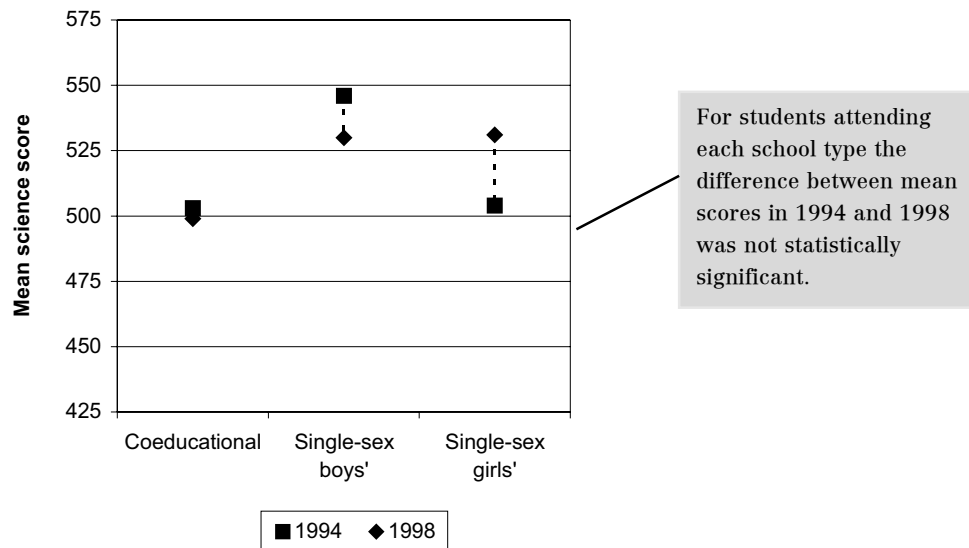
<sup>6</sup> Most students within each type of school (at least two-thirds) fell into the medium HER category.

analysed by school type, no statistically significant differences (after adjusting for multiple comparisons of means) in mean science scores between school type emerged. By contrast, students categorised as high on the HER Index have statistically significantly higher mean science scores than those categorised as being at the medium level of the HER Index within the same school type.

In summary, it does not appear that school-type alone should be considered as having the major influence on science achievement, despite the association between the two.

There were no statistically significant differences between the mean science scores achieved by students in each school type in TIMSS-98 and their counterparts in 1994. Figure 4.7 plots the mean science achievement scores of students in each of the three school types for 1994 and 1998 (see Table C.12 in Appendix C for further details).

**FIGURE 4.7: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY SCHOOL TYPE**



## THE SCIENCE CONTENT REPORTING CATEGORIES

As with TIMSS-94, the TIMSS-98 science items were based around discrete content areas. However, the content areas differed slightly between TIMSS-94 and TIMSS-98 and, unlike the main science scale, the sub-scales between the two studies are not directly comparable. The six TIMSS-98 science content areas are as follows:

1. *Earth Science* — includes earth features, earth processes, and the universe
2. *Life Science* — includes diversity, organisation and the structure of living things, life processes and systems enabling life functions, life spirals, genetic continuity and diversity, human biology and health
3. *Physics* — includes energy types, sources, and conversion, light heat and temperature, physical transformations, physical properties and structure of matter, and forces and motion.
4. *Chemistry* — includes classification and structure of matter, physical and chemical properties, and chemical transformations

5. *Environment and Resource Issues* — includes pollution, conservation of land, water, and sea resources, conservation of material and energy resources and effects of natural disasters
6. *Scientific Inquiry and the Nature of Science* — includes the nature of scientific knowledge, the scientific enterprise, and the interactions of science, technology and society.

For further details see McKnight et al, 1993; Robitaille et al, 1993; and Martin et al, 2000.

### Test Curriculum Matching Analysis

There was a total of 143 science items in TIMSS-98. These were variously formatted as multiple-choice, dichotomous (single mark) open-ended and polychotomous (multiple mark) open-ended response items.

About one-third of the items were 'trend' items — that is, they were common to both TIMSS-94 and TIMSS-98. The remaining two-thirds were new items, replacing those TIMSS-94 items that were released into the public domain in 1997. The new items were developed to maximise the 'fit' of the test to curricula across countries.

As was the case in TIMSS-94, a Test Curriculum Matching Analysis was carried out in order to determine which items, within the final pool of items, were explicitly in the *intended* curriculum for each country. Table 4.9 provides a breakdown by science content area.

**TABLE 4.9: PROPORTION OF TIMSS-98 SCIENCE ITEMS JUDGED TO BE IN THE NEW ZEALAND INTENDED CURRICULUM FOR YEAR 9 STUDENTS, BY CONTENT AREA**

Content reporting category	Number of items *	Score points <sup>^</sup> allocated to each content category	Items in the <i>intended</i> curriculum (%)
Earth Science	22	23	64
Life Science	39	42	64
Physics	39	39	90
Chemistry	19	22	79
Environment & Resource Issues	12	14	83
Scientific Inquiry & the Nature of Science	12	13	92

Notes:

\* Number of items in each reporting category does not reflect the fact that there were different parts to the items. For the Test Curriculum Matching Analysis, the entire item was examined when judging whether or not it was in the intended curriculum for NZ Year 9 students.

<sup>^</sup> Some free-response items were divided into two or three parts; each part of an item was treated as a single item for scoring purposes.

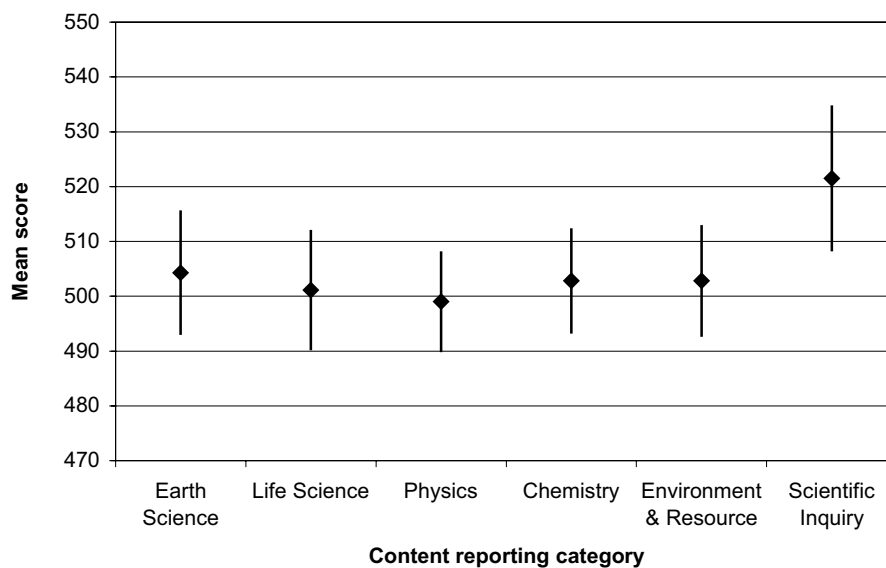
Using Levels 4 and 5, and in some cases Level 3, of *Science in the New Zealand Curriculum* (Ministry of Education, 1993), the content of 77 percent of the science items was judged to be in the *intended* curriculum, for the majority of Year 9 students (see Table 4.9 for details). Note that this percentage does not directly translate into the total raw score points available to students, as polychotomous open-ended items (ie, where partial credit may be awarded) are counted here as a single item.

As stated in Chapter 3 it is important to remember that although the content of an item may have been explicitly stated in the curriculum guide it does not necessarily mean that all students have been exposed to or taught that content. Conversely, students may be exposed to science content that would be considered peripheral or even outside of the intended science curriculum.

### ACHIEVEMENT IN THE SCIENCE CONTENT REPORTING CATEGORIES

The six content areas form independent sub-scales — they should not be compared directly with the overall science scale. However, as all the sub-scales are placed on a common reporting scale (or metric) it is possible to compare the relative strengths and weaknesses of populations with regard to the different content areas. Note that the metric was set so that the international mean was placed at 488 for each of the science content areas. In all content areas New Zealand students achieved a mean score higher than the international mean; but only in *Scientific Inquiry and the Nature of Science* was the difference statistically significant (after adjusting for multiple comparisons) (Figure 4.8).

**FIGURE 4.8 YEAR 9 STUDENTS' MEAN SCORES IN EACH SCIENCE REPORTING CATEGORY IN 1998**



The data points are the mean scores. The vertical lines extending from the data point show the 95% confidence interval around the mean, ie,  $\pm 2$  standard errors.

From Figure 4.8 it can be seen that *Scientific Inquiry and the Nature of Science* is clearly an area of strength for New Zealand's Year 9 students. However, when examining the relative content area performance of any population in an international context two (related) factors must be considered. Firstly, differences in curriculum emphases between countries must be taken into account (for more on this see Martin et al, 2000) and, secondly, differences between a country's curriculum and the items in the test instruments must be considered. For example, at first glance, it is not surprising that New

Zealand Year 9 students' performance in *Earth Science* was weak relative to their performance in *Scientific Inquiry and the Nature of Science*.

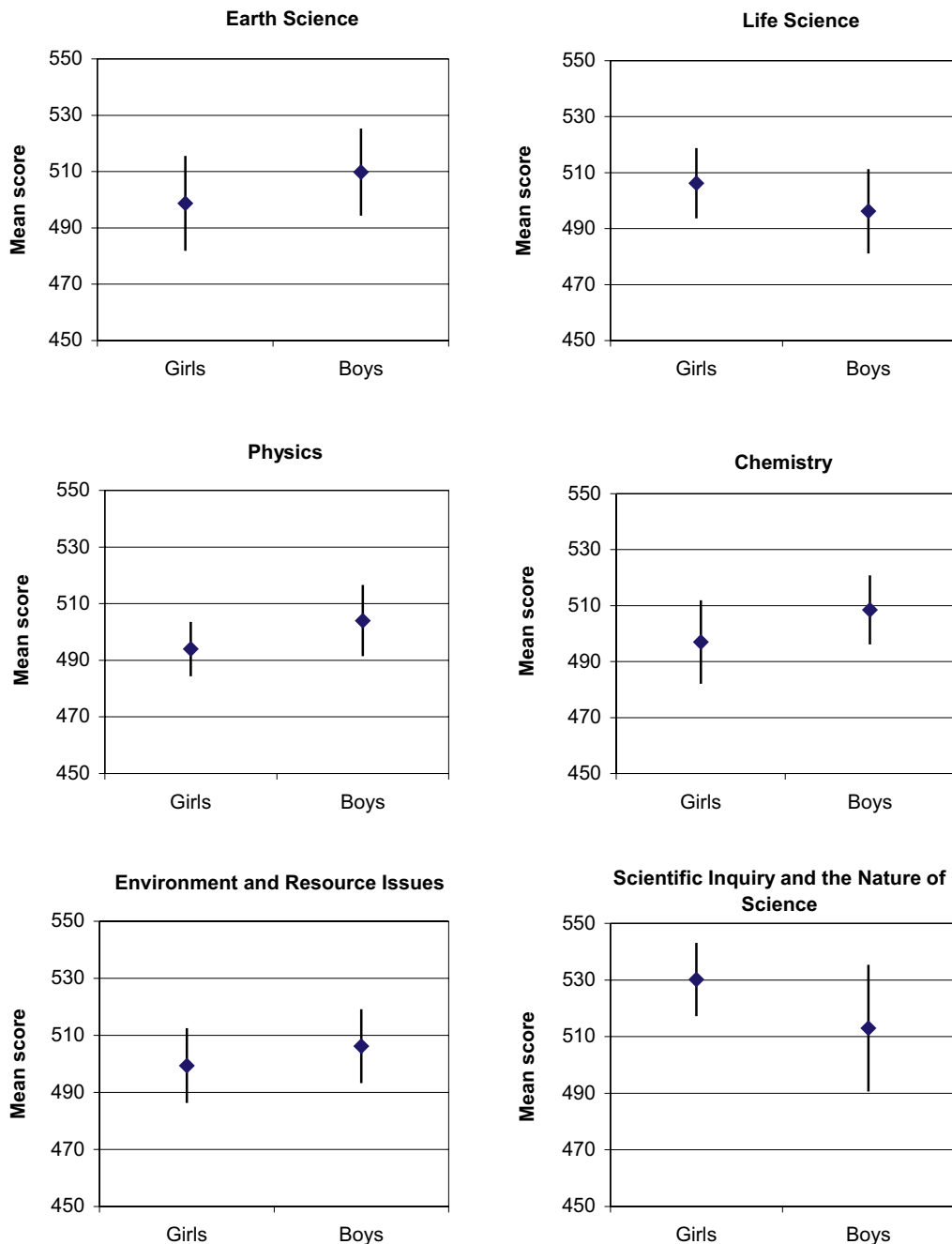
It may reasonably be speculated that this was due, in part, to the fact that only two-thirds of the *Earth Science* test items were in the *intended* curriculum compared with over 90 percent of the *Scientific Inquiry and the Nature of Science* items. But the picture is more complex than this because 90 percent of the *Physics* items were also found to be in the *intended* curriculum yet the relative performance of New Zealand students in this area was clearly weaker than in *Scientific Inquiry and the Nature of Science*. A possible explanation may be found by examining the relative emphasis that other countries in the study placed on these two areas — for example, if most other countries place considerably more emphasis on *Physics* than *Scientific Inquiry and the Nature of Science* it may account for why New Zealand students performed well on the latter, relative to their international counterparts. Such examination is, however, beyond the scope of the current report.

### **Gender and achievement in the science content reporting categories**

In New Zealand, there were no significant differences found between the mean score of boys and the mean score of girls on any of the six science content area sub-scales outlined previously (see Figure 4.9 and Table C.13 in Appendix C for detail of results).

New Zealand boys achieved a mean score above the international mean for boys in every science content area but none of these differences was statistically significant (significance tests were adjusted for multiple comparisons, see TN.5 in Technical Notes).

New Zealand's girls achieved mean scores above the international means for girls in each of the six science content reporting areas but only two of these differences were statistically significant (after significance tests were adjusted for multiple comparisons). These were *Physics* and *Scientific Inquiry and the Nature of Science*. For *Physics* the international mean for girls was 477 whereas the mean for New Zealand girls was 494. The corresponding mean scores for *Scientific Inquiry and the Nature of Science* were 489 (international girls) and 530 (for New Zealand girls).

**FIGURE 4.9: YEAR 9 STUDENTS' MEAN SCORES FOR THE SIX SCIENCE CONTENT REPORTING CATEGORIES AREAS IN 1998, BY GENDER**

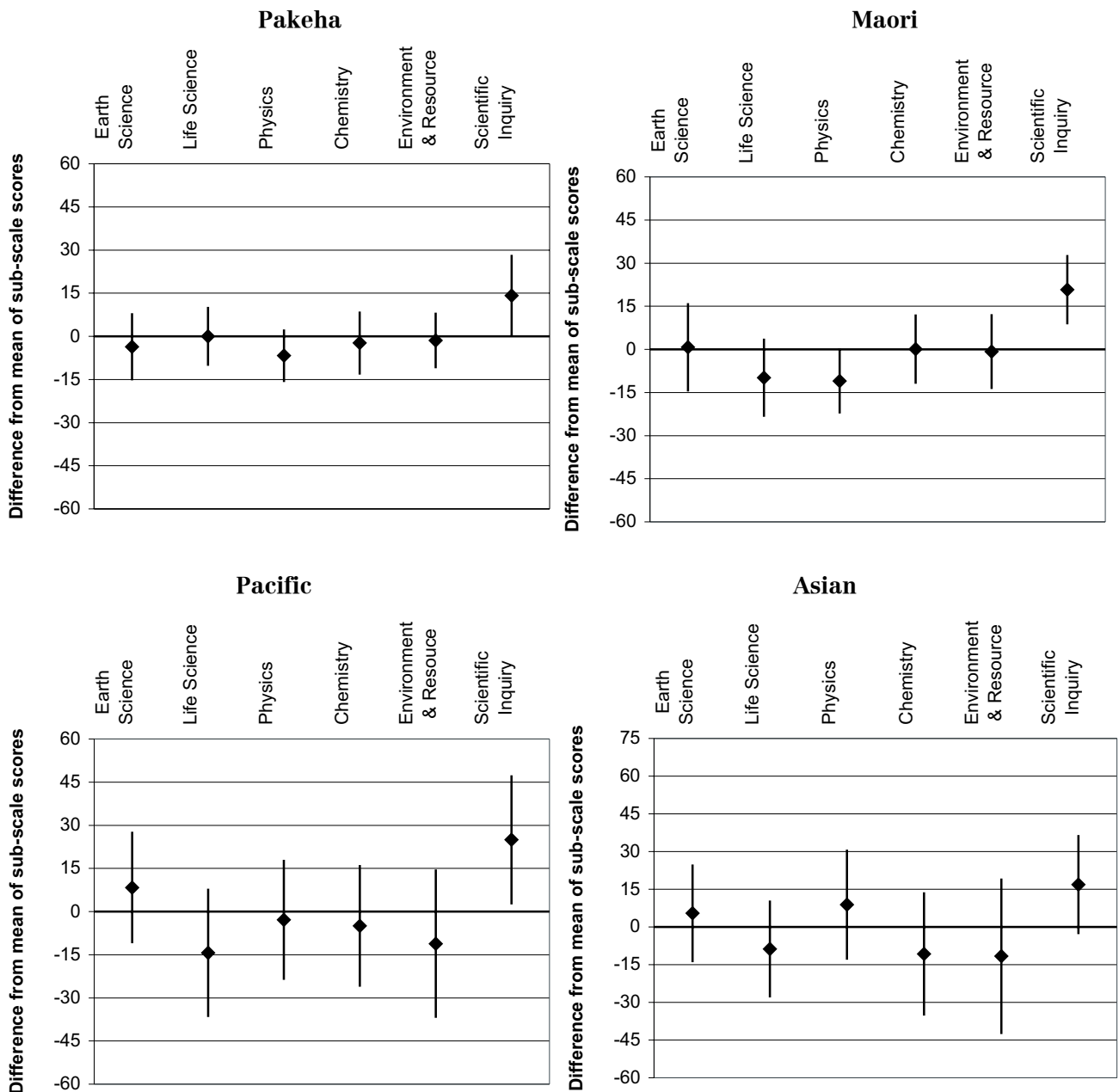
The data points are the mean scores. The vertical lines extending from the data point show the 95% confidence interval around the mean, ie,  $\pm 2$  standard errors.

### Ethnicity and achievement in science content areas

The patterns of relative achievement between ethnic groupings in the content area sub-scales generally reflect the pattern for the overall science scale, although there were some notable differences. On the one hand, for example, Pakeha/European students achieved significantly higher means than Maori and Pacific students in all six science content areas. On the other hand, Maori significantly outperformed Pacific students only in *Chemistry*; and only in *Earth Science* and *Physics* did Asian significantly outperform Maori. Details of the mean sub-scale scores achieved by each ethnic grouping can be found in Table C.14 in Appendix C.

It was possible to also examine each ethnic grouping's strengths and weaknesses across the different content areas. Figure 4.10 plots the content area sub-scale scores achieved by each ethnic group, with the zero-point representing the average of the sub-scale scores *for that group*. For the four groups, the patterns of strength and weakness were, on the whole, similar. For example, *Nature of Science and Scientific Inquiry* was clearly an area of relative strength for all groups. The similarities in patterns of achievement may reflect the relative emphases in the *intended* and *implemented* curriculum.

**FIGURE 4.10: RELATIVE STRENGTHS AND WEAKNESSES IN THE SCIENCE CONTENT REPORTING CATEGORIES IN 1998, BY ETHNIC GROUPING**



Each data point represents the difference from the mean of the sub-scale scores. The vertical lines extending from the data point show the 95% confidence interval around this point, ie,  $\pm 2$  standard errors.

*Chapter 4 has focused on the trends in New Zealand Year 9 students' achievement in science — Chapter 5 presents contextual information for these students.*

