CHAPTER 2: Mathematics Achievement

KEY POINTS

- There was a small non-statistically significant increase in the mean mathematics achievement of Year 5 students from 1994 to 1998.
- Year 5 boys in 1998 on average achieved significantly higher mathematics scores than their male counterparts did in 1994.
- The most notable improvement in Year 5 boys' mean mathematics performance from 1994 to 1998 was observed for Maori boys which was of statistical significance.
- Year 5 girls in 1998 on average achieved at about the same level as their female counterparts did in 1994.
- There was a marked decrease in the difference between mean scores of Year 5 students who reported speaking English frequently in the home and those who rarely did, from 1994 to 1998.
- *Whole Numbers* was a content area of relative weakness, while *Data Representation, Analysis and Probability* was an area of relative strength in both 1994 and 1998.

This chapter focuses on the overall mathematics performance of the Year 5 (standard 3) students who participated in TIMSS-98, compared with that of Year 5 (standard 3) students who participated in TIMSS-94[1]. The latter part of this chapter summarises results for a selection of items used in both the TIMSS-94 and TIMSS-98 assessments. Since New Zealand was the only country to administer TIMSS-98 at the middle primary level, there are no international results for inclusion in this or subsequent chapters.

REPORTING STUDENT ACHIEVEMENT RESULTS

As described in Chapter 1, three approaches are used in this report to describe student achievement:

- Item Response Theory (IRT) scale scores to describe overall achievement;
- Mean percent correct scores to examine student achievement on *trend* (items kept secured after TIMSS-94) and *released* (test items released into the public domain after TIMSS-94) items;
- Percentage of students who answered each individual mathematics item correctly.

Standard errors (se) have been calculated using the Jackknife Repeated Replication procedure. A 95 percent confidence interval applies. See Technical Notes TN.1 and TN.2 in Appendix E for more detail on the statistical methods used.

TRENDS IN MATHEMATICS ACHIEVEMENT

As noted in Chapter 1, two sets of information are used to examine trends in Year 5 students' achievement. Firstly, the overall 1998 mathematics achievement results, in the form of IRT scale scores, are compared with the IRT scale scores for 1994. Secondly, there is some discussion on the results for the *trend* item set. *Trend* refers to the group of questions that remained secure for use in future studies, while the *released* test items, representing about two-thirds of the test item pool were placed into the public domain after the conclusion of TIMSS-94. This then, is the principal reason that analyses of *trend* items have been reported — *trend* results substantiate overall findings when they are consistent and place caveats around overall findings where they are inconsistent.

OVERALL MATHEMATICS ACHIEVEMENT

Table 2.1 presents the mean scores for Year 5 students. Standard errors (se) appear in parentheses, and have been calculated using the jackknife repeated replication procedure[2]. We can say, with 95 percent confidence that the true mean lies within two standard errors of the sample (reported) mean.

TABLE 2.1 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1994 AND 1998

Year of assessment	Overall mathematics mean (se)
1994	469 (4.4)
1998	481 (5.6)

Note: (se) Standard errors appear in parentheses.

[1] *As noted in Chapter 1, TIMSS-94 involved middle-primary students at two class levels — standards 2 and 3 (or international grades 3 and 4). Years 4 and 5 students are mostly in these two class levels. TIMSS-98 assessed students at a class level equivalent to standard 3 level and involved mostly Year 5 students. The students involved in TIMSS-98 will hereafter be referred to as 'Year 5' students.*

[2] *This approach was necessary because the design of the study incorporated multiple-matrix-sampling; that is, each student received only a sample of the test questions.*

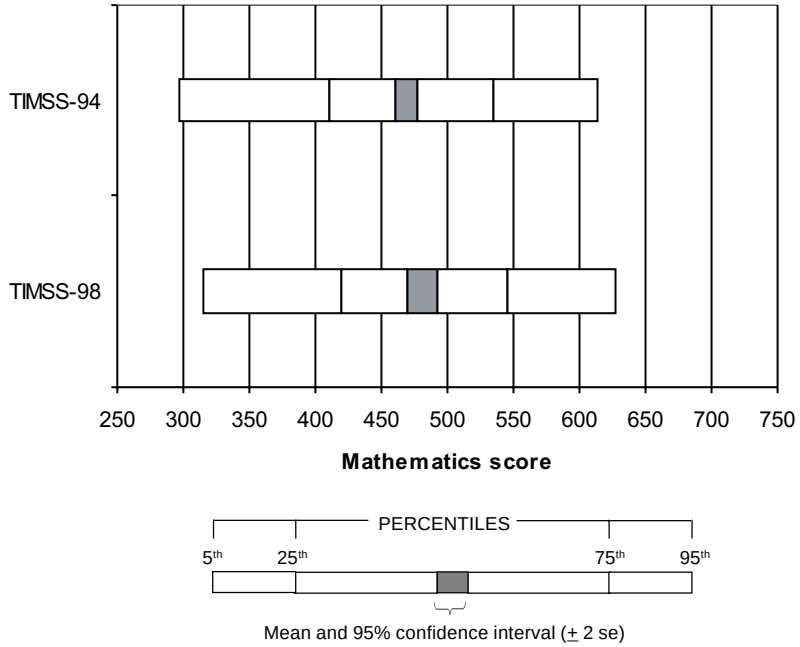
As shown in the table there was a small increase in mean mathematics achievement over the four-year period from 1994 to 1998. However, the difference was found not to be of statistical significance.

Figure 2.1 contains information on the distribution of the mathematics scores for Year 5 students for both 1994 and 1998.

Interpretation of the percentiles

The percentages of students performing below or above particular points on the scale are shown for each year. These points represent the outer limits of achievement. The lowest outer limit is the 5th percentile — the score at which only five percent of students achieved a lower score, and 95 percent of students achieved a higher score. The highest outer limit is the 95th percentile — the score at which only five percent of students achieved a higher score, and 95 percent of students a lower score. Ninety percent of the Year 5 student scores were between the 5th and 95th percentiles.

FIGURE 2.1 DISTRIBUTION OF YEAR 5 STUDENTS' MATHEMATICS SCORES FOR 1994 AND 1998



While the range of scores was about the same in both years (ie, the difference between the 5th and the 95th percentiles), the lower achieving students achieved slightly higher scores in TIMSS-98 than was the case four years earlier; so too did the higher achieving students. That is, the 5th percentile (five percent of students achieved below this score point) in 1998 was 315 compared with 297 in 1994, while the 95th percentile was 627 compared with 613. See Table B.1 in Appendix B for details of the percentiles for 1994 and 1998.

Trend item results

Year 5 students achieved on average a higher score on the *trend* item set in 1998 compared to their Year 5 student counterparts in 1994. However the difference was found not to be of statistical significance (see Table B.2 in Appendix B). This finding would suggest that the small increase in the overall mathematics achievement is, in part, attributed to the improvement on the *released* items that was found to be of statistical significance ($\alpha = 0.05$).

Gender

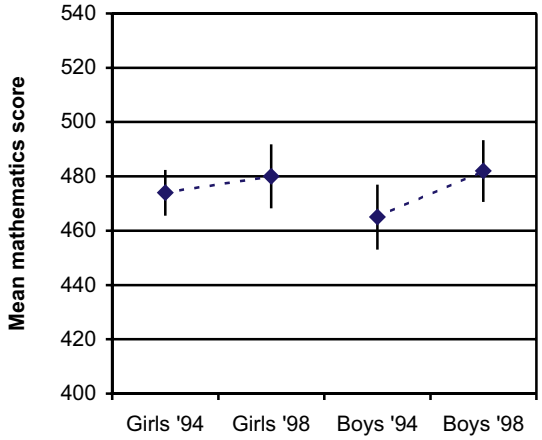
Overall results

In 1998 boys, on average, achieved scores higher than their male counterparts in 1994 with the difference between mean scores (17 scale score points) of statistical significance (α=0.05). By comparison, girls on average achieved at about the same level in 1998 as their female counterparts in 1994 with a mean difference (six scale score points) that did not reach statistical significance. How big was the difference for boys? Effect sizes have been used to find out the size or magnitude of the difference between the means for 1994 and 1998 (see Technical Note TN.5 in Appendix E). The effect size (d=0.18) represents a relatively small increase in mean achievement from 1994 to 1998.

Whereas girls in 1994, on average, achieved slightly higher (non-significant) mathematics scores than boys, the mean scores for boys and girls were about the same in 1998. While 'overall' gender scores can sometimes mask changes within specific sub-groups, this information is important, given the concerns expressed by commentators in recent years over the performance of boys (eg Alton-Lee & Praat, 2000).

Figure 2.2 presents the mean scores for boys and girls for 1994 and 1998 (also see Table B.3 in Appendix B for details). It is also important to remember that those boys and girls are not homogeneous groups. Further analyses were carried out to examine the performance of boys and girls within each of the main ethnic groupings, the results of which are presented later in this chapter.

FIGURE 2.2 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1994 AND 1998, BY GENDER



Notes: *See Chapter 4 for the proportions of students who were boys or girls in 1994 and 1998.*

The data points are the mean scores. The vertical lines extending from the data points show the 5% confidence interval around the mean, ie ± 2 standard errors.

Trend item results

The apparent improvement in Year 5 boys' performance can be attributed to significant improvements on both the trend and the released test items (of 3 and 4 percentage points respectively). There was virtually no change in Year 5 girls' mathematics achievement on either the trend or the released items between 1994 and 1998 (see Table B.4 in Appendix B for details).

ETHNICITY[3]

Overall results

A positive feature of the results for two of the ethnic groupings — Asian and, to a lesser extent, Maori — was some improvement in mathematics performance between 1994 and 1998 (see Table 2.2). However, the differences in mean scores for these groupings were found not to be of statistical significance. There was less variation in the mean scores for the Pakeha/European and Pacific groupings over the same period.

TABLE 2.2 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1994 AND 1998, BY ETHNIC GROUPING

Year of assessment	Mean mathematics score (se)				
	Pakeha/ European	Maori	Pacific	Asian	Other
1994	493 (3.9)	427 (8.2)	412 (11.0)	483 (16.9)	475 (15.1)
1998	502 (5.0)	445 (7.3)	416 (15.1)	516 (9.9)	481 (15.0)

Notes: *See Chapter 4 for the proportions of students in each ethnic grouping for 1994 and 1998.*

(se) Standard errors appear in parentheses.

In 1994, Pakeha/European students, on average, achieved mathematics scores slightly higher than their Asian counterparts. This situation was reversed four years later. However, in both years the difference between mean scores for these two groupings was not of statistical significance.

As was the case in 1994 and 1998, there continued to be a significant differential between the mean performance of Year 5 students in both the Maori and Pacific groupings and their counterparts in the Asian and Pakeha/European groupings. Tables B.5a and B.5b in Appendix B show the effect sizes between each grouping for both 1994 and 1998. (See Technical Note TN.5 in Appendix E for more details on effect sizes.)

Trend item results

Non-statistically significant increases in mean mathematics achievement (2 to 5 percentage points) from 1994 to 1998 were also observed on the trend items for three out of the four main ethnic groupings. No change was observed for Pacific. All ethnic groupings reported improvement on the released item set ranging from three to eight percentage points (for Pacific and Asian respectively). This suggests that any improvement observed in overall mathematics achievement for all ethnic groupings was due mainly to small increases in mean achievement on the released item set (albeit statistically significant for just Pakeha/European) as previously noted. See Tables B.6a and B.6b in Appendix B for details.

GENDER AND ETHNICITY

Overall results

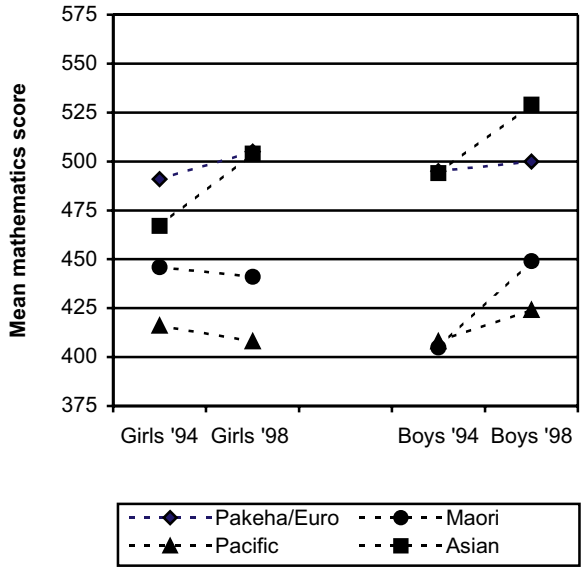
When achievement for ethnic grouping was examined by gender, the largest increase in mean mathematics scores between 1994 and 1998 was observed for Maori boys (44 scale score points, α=0.05),

[3] *Statistics New Zealand hierarchical classification of ethnicity was used for these analyses. See Chapter 4 for more details. Those students in the "Other ethnic groups" were excluded from most of the discussion.*

and to a lesser extent, Asian girls and boys (37 and 35 scale score points respectively). In comparison, the sub-groups with effectively no change over the same period included Pakeha/ European boys, and Maori and Pacific girls, while Pakeha/European girls' and Pacific boys' increases were not large enough to be of statistical significance.

In 1994, Maori girls significantly outperformed their male counterparts by an average of 41 scale score points (α=0.05). However four years later, the differential favouring Maori girls had all but disappeared (8 scale score points higher for boys). The mean scores are represented in Figure 2.3, while the actual figures and standard errors are reported in Table B.7 in Appendix B.

FIGURE 2.3 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1994 AND 1998, BY GENDER AND ETHNIC GROUPING



Trend item results

The increase in the overall achievement was also reflected in the mean percent correct scores for the *trend* items. The difference between Maori boys' mean percent correct scores from 1994 to 1998 was seven percentage points, which was found to be statistically significant increase (α=0.05). Maori girls in 1998 on average achieved scores about the same as their 1994 counterparts.

The "disappearance" of the overall differential that was observed between Maori girls and Maori boys in 1994 (5 percentage points in favour of girls) was also observed on the *trend* item set. Relatively large increases in mean percent correct scores were also observed for Asian girls and Asian boys, however, due to the relatively large standard errors (ie, due in part to the relatively small sample sizes) the differences were not found to be of statistical significance. (See Table B.8 in Appendix B for details.)

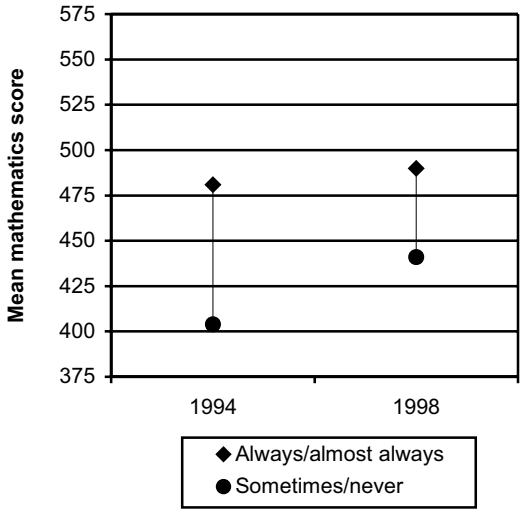
HOME LANGUAGE

Overall results

Earlier IEA studies on which Garden (1984), Lamb (1987), Wagemaker (1993), and Chamberlain M. (1997) have reported, show that when the predominant home language differs from that used in the classroom, the level of achievement is likely to be lower. Year 5 students were asked to indicate the

extent to which they spoke English at home[4] and on the basis of their responses were divided into two groupings. Eighty-five percent of students reported that they always or almost always spoke English at home in 1998 (compared with 87% in 1994), and 15 percent reported that they only sometimes or never spoke English (13% in 1994). The mean mathematics scores of these two language groupings for both 1994 and 1998 are presented in Figure 2.4 (for details of the actual figures see Table B.9 in Appendix B).

FIGURE 2.4 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1994 AND 1998, BY THE DEGREE THAT ENGLISH IS SPOKEN IN THE HOME



In 1994, the difference in mean mathematics achievement between those who rarely and those who usually speak English at home was 77 scale score points. By 1998, the difference between mean scores for these two groups had reduced by approximately one-third to 49 scale score points. This reduction is largely attributed to the increase in mean achievement of almost 40 scale scores points for students who reported only 'sometimes/never' speaking English in their home. (Note: the difference in mean scores between 1994 and 1998 for students in this grouping was statistically significant α=0.05). The decrease in the differential in mean achievement between students in the two language groupings is also illustrated by a shift in the magnitude of the difference or effect size (see Technical Note TN.5 in Appendix E). In 1994, for mathematics, the effect size between the categories of students was d=0.83 and in 1998 it was d=0.53.

In 1998, Year 5 students in the Pakeha/European and Maori groupings were more likely to report that they usually spoke English in the home than students in the Pacific and Asian groupings. Ninety-five percent and 81 percent of students in the two former groupings reported that they usually spoke English compared with just under one-half of the Pacific students and one-third of Asian students.

Figure 2.5 presents the mean mathematics scores for students in the two language categories by ethnic grouping. For details of the proportions of students in each category and the mean scores and standard errors for both 1994 and 1998, refer to Tables B.10 and B.11 in Appendix B.

[4] *Only Year 5 students in English-medium classes were assessed in both TIMSS-94 and TIMSS-98.*

FIGURE 2.5 YEAR 5 STUDENTS' MEAN MATHEMATICS SCORES FOR 1998, BY THE DEGREE THAT ENGLISH IS SPOKEN IN THE HOME AND ETHNIC GROUPING

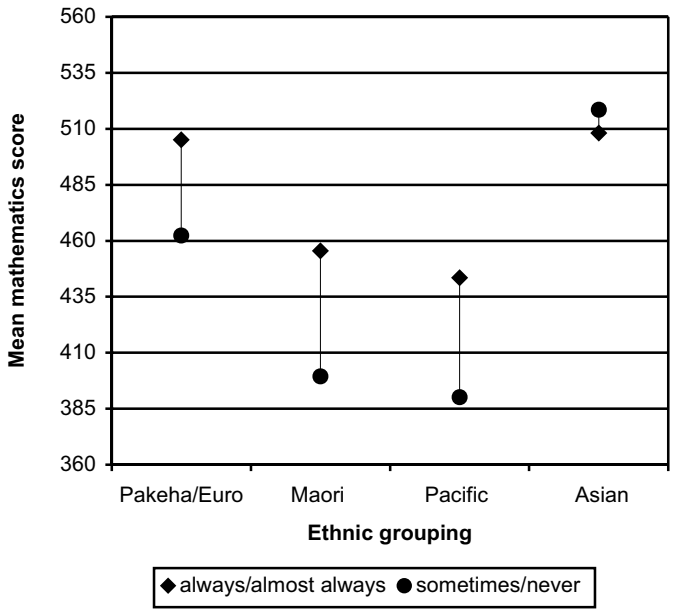


Figure 2.5 shows that for three out of the four groupings — the exception being Asian — the mean achievement of students who rarely spoke English at home was lower than that of their respective counterparts who spoke English more frequently.

There was a statistically significant increase in mean mathematics scores for both Asian and Pakeha/European Year 5 students who rarely spoke English at home, over the four years from 1994 to 1998 (Asian = 50 scale scores points, α=0.05; Pakeha/European = 40 scale score points, α=0.05). There were smaller non-significant increases in mean achievement for Maori and Pacific students (22 and 13 scale score points respectively) who rarely spoke English in the home.

Trend item results

The main point of interest on the *trend* mathematics item results for Year 5 students who rarely spoke English at home was an eight point increase in mean percent correct between 1994 and 1998 (α=0.05). There was a similar improvement (7 percentage points) on the *released* items over the same time period (α=0.05). Further information on the results is available in Table B.12 in Appendix B.

MATHEMATICS CONTENT REPORTING CATEGORIES

Overall results

This section summarises TIMSS-98 results in relation to the mathematics reporting categories of *Whole Numbers; Fractions and Proportionality; Geometry and Measurement*; and *Data Representation, Analysis, and Probability*. In some instances comparisons are made with information collected in 1994. Readers should note that there is no discussion on how students achieved on the *trend* items in each of the reporting categories due to the small number of items in some categories. Any change in performance should therefore be noted with caution.

The scores for each of the reporting categories were calculated separately from the overall mathematics scale score and therefore should be regarded as independent measures. That is, the means for content reporting category cannot be compared to the overall mean reported at the beginning of this chapter.

It is also worth recalling from Chapter 1 that within each content domain, students were required to demonstrate a range of performances in order to answer the individual questions correctly. For example, approximately one-quarter of the *Data Representation, Analysis, and Probability* items required students to demonstrate *solving* skills to answer correctly, while a similar proportion (26%) of items in *Whole Numbers* were categorised as *routine* (for more detail see Adams & Gonzalez, 1996; Garden, 1997).

Figures 2.6a and 2.6b (see also Table B.13 in Appendix B) present the mean scores in each content reporting category for 1994 and 1998 respectively.

FIGURE 2.6A YEAR 5 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS REPORTING CATEGORY IN 1994

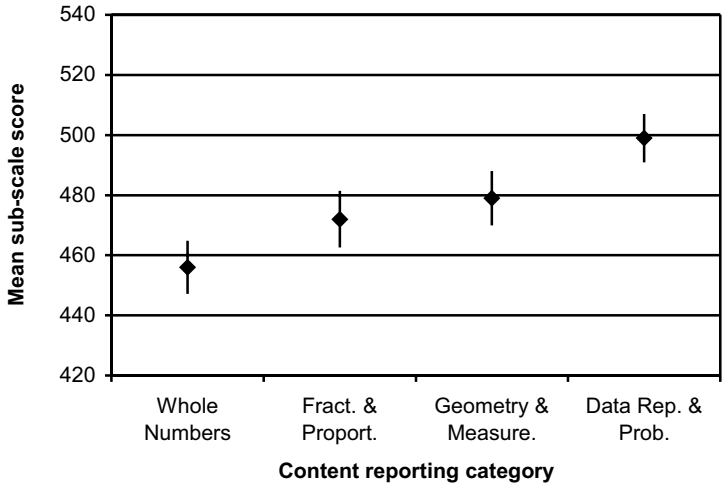
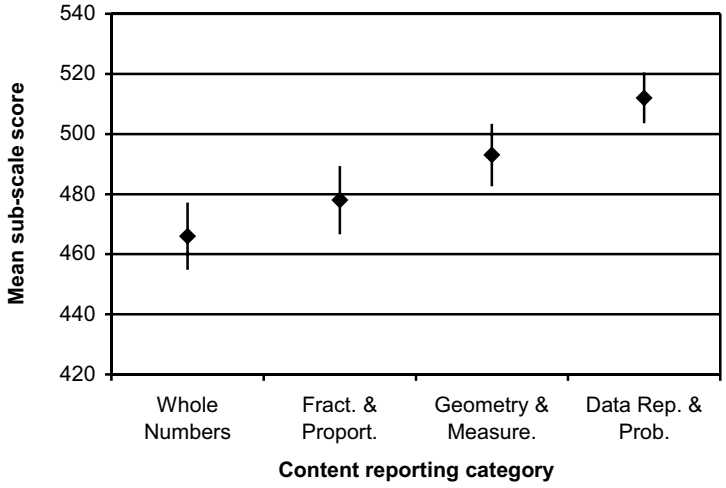


FIGURE 2.6B YEAR 5 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS REPORTING CATEGORY IN 1998



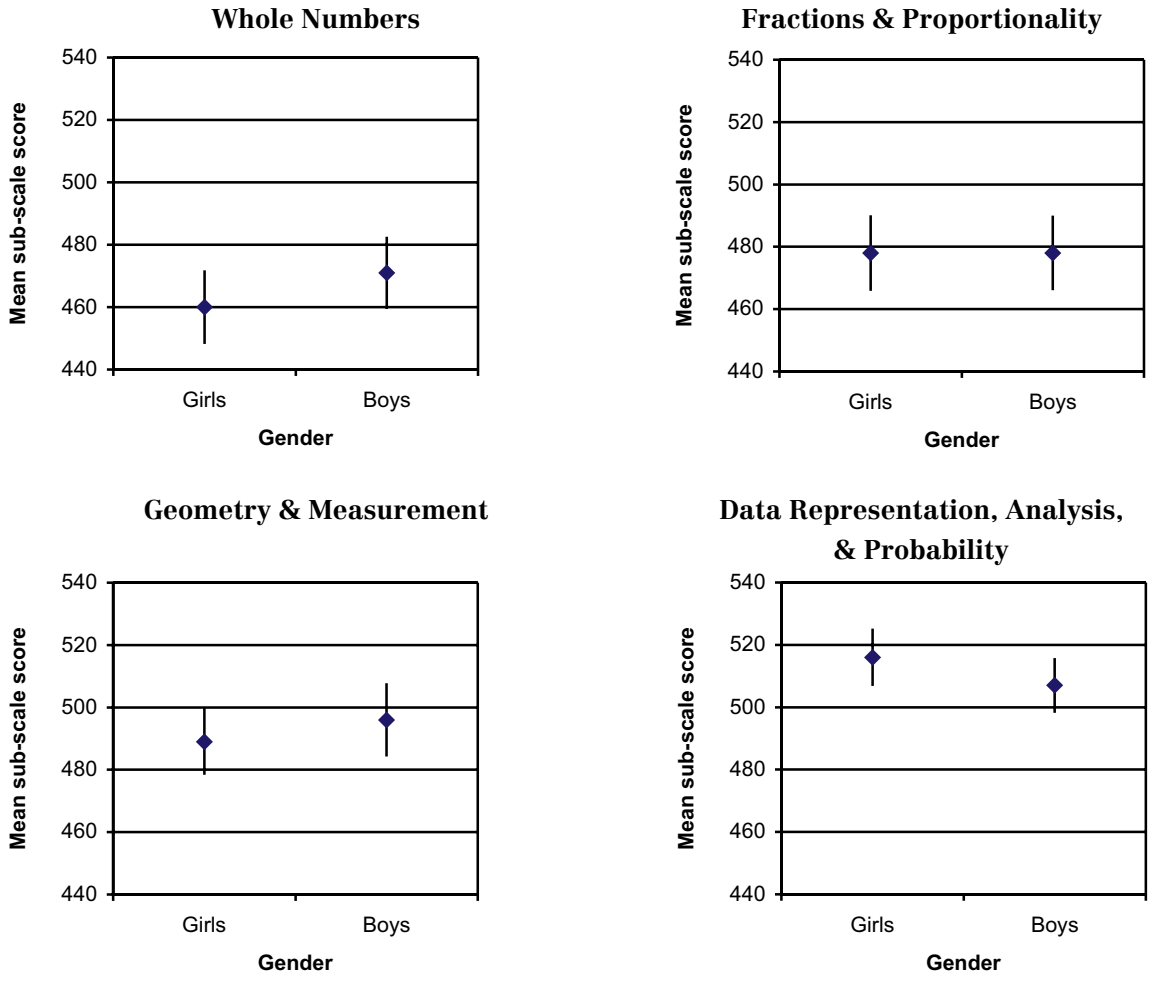
The data points are the mean scores. The vertical lines extending from the data points show the 95% confidence interval around the mean, ie ± 2 standard errors.

When Year 5 student achievement was examined for each of the reporting categories, there were small increases in all of the means for the categories between 1994 and 1998. The increases ranged from six scale score points observed for *Fractions and Proportionality*, through to 13 scale score points for *Geometry and Measurement*, and *Data Representation, Analysis, and Probability*.

In 1994, the mathematics content area that was area of relative strength for Year 5 students was *Data Representation, Analysis, and Probability*. This was also the case four years later in 1998. And, although the mean for Whole Numbers increased by 10 scale score points from 456 in 1994 to 466 in 1998, the area was still found to be an area of relative weakness for Year students.

Figure 2.7 provides a graphical representation of mean achievement for each mathematics reporting category for 1998 by gender. For details about the actual figures see Tables B.14 in Appendix B.

FIGURE 2.7 YEAR 5 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS REPORTING CATEGORY IN 1998, BY GENDER



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

The overall increase in the mean mathematics score observed for boys was also reflected, with just one exception (Fractions and Proportionality), in each of the content reporting categories. The increases over the four-year period from 1994 to 1998 were of statistical significance (α=0.05). There was virtually no change in girls' mean achievement in any reporting area.

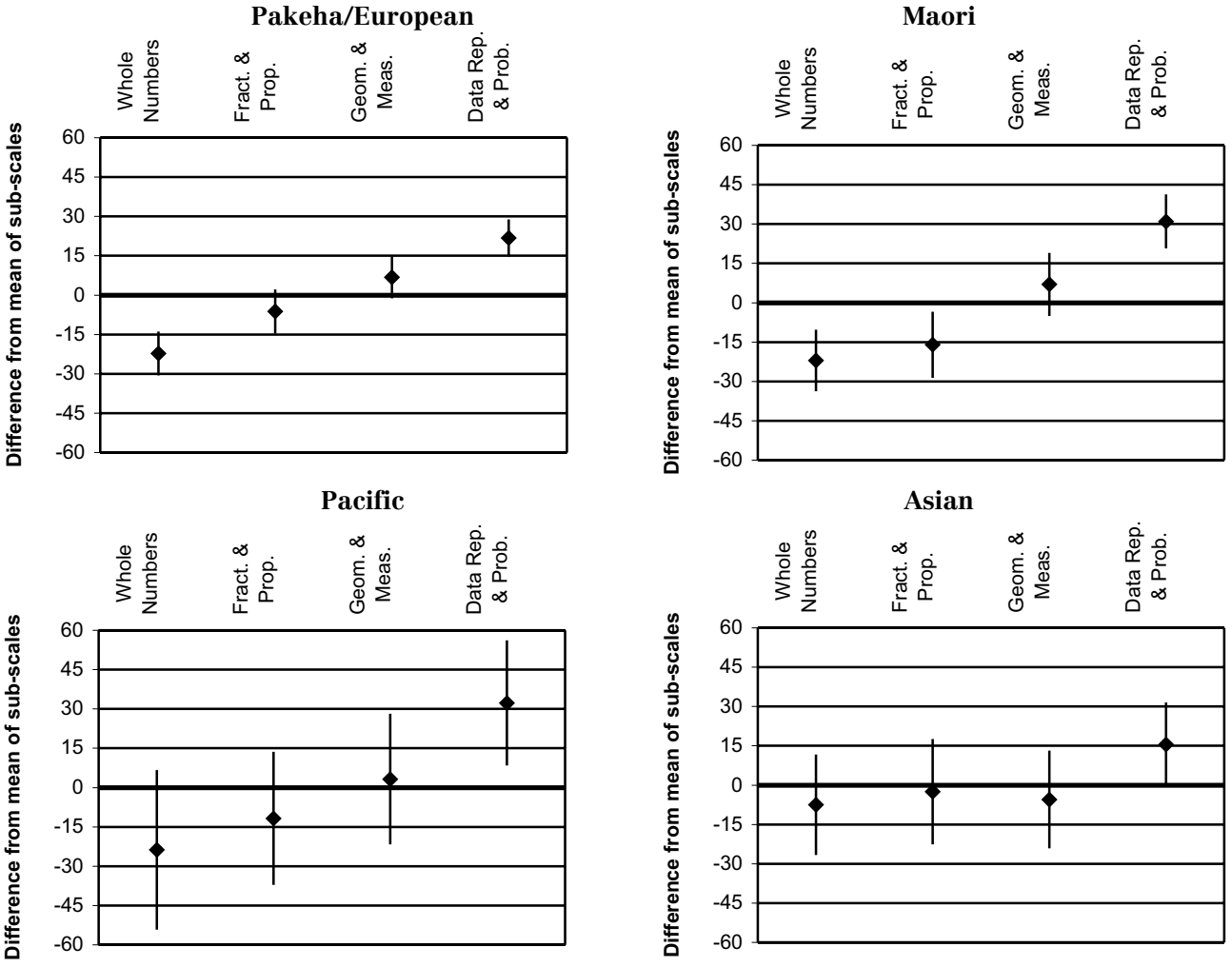
Whereas girls' mean scores in 1994 were consistently higher than boys' mean scores in all reporting categories, including a statistically significant difference (α=0.05) for Data Representation, Analysis, and Probability, this was not the case in 1998. Although not of statistical significance, boys' mean achievement in 1998 in *Whole Numbers* and *Geometry and Measurement* surpassed that of their female counterparts, and equalled them in *Fractions and Proportionality*. Only in *Data Representation, Analysis, and Probability* did girls, on average, achieve higher scores than boys. However, the difference between the two groups of students favouring girls reduced from 22 scale score points in 1994 to just nine scale score points in 1998.

As noted on page 19, the content reporting category scores cannot be compared to the overall mathematics score. However, in order to gauge areas of relative weakness or strength, it is possible to compare the mean for a given reporting category with the mean calculated for the sub-scales.

Comparing the performance across the mathematics content areas by gender in 1998, both boys and girls were relatively weak in *Whole Numbers* while *Data Representation, Analysis, and Probability* was an area of strength relative to their overall performance.

Figure 2.8 provides a graphical representation of the mean achievement in each mathematics reporting category for 1998 by ethnic grouping. For details about the actual figures see Table B.15 in Appendix B.

FIGURE 2.8 YEAR 5 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS REPORTING CATEGORY IN 1998, BY ETHNIC GROUPING



The data points are the mean subscale scores. They are plotted by the difference from the mean of the four content area scores (represented by 0). The vertical lines extending from the data points show the 95% confidence interval around the score, ie ± 2 standard errors.

On average in 1998, the achievement of students in each content reporting category area by ethnic grouping mirrored the achievements of these students overall. That is, on average Asian and Pakeha/European students achieved higher scores in each reporting category than Maori and Pacific students. However, there was some variation from 1994 to 1998 across the ethnic groupings.

The overall relatively large increases in the mean mathematics scores observed for Maori and Asian students were also reflected in increases in each of content reporting categories. The largest increases for Maori were observed in *Geometry and Measurement* (26 scale scores points) and *Data Representation, Analysis, and Probability* (21 scale scores points) with the increases of statistical significance (α=0.05). While the increases in each reporting category for Asian were numerically larger (eg by 10 or more scale score points) than those for Maori, they were found not to be of statistical significance due in part to the very large standard errors (partly because of the small sample size) for 1994.

Data Representation, Analysis, and Probability was found to be area of relative strength for all four ethnic groupings. *Whole Numbers* was an area of relative weakness for both Maori and Pakeha/European students and to a lesser extent (due in part to the larger standard errors), Pacific students.

PERFORMANCE ON A SELECTION OF MATHEMATICS ITEMS

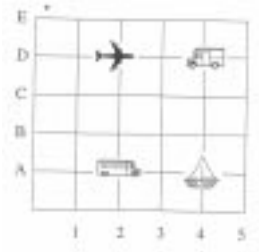
This final section of the mathematics results will focus on how Year 5 students responded to a selection of the individual *released* test questions. A (weighted) percentage is used to describe the proportion of students answering correctly in both TIMSS-94 and TIMSS-98. Note that these items are reported as examples only. Systematic differences are more easily captured in the overall scale and content area sub-scale scores.

An example of one of the easiest mathematics items for Year 5 students in 1994 and 1998.

Example 1

Content area: *Geometry (knowing)*

L3. This is a game board.



Which object is located at (2,D)?

A. The plane 

B. The truck 

C. The bus 

D. The boat 

1994 Year 5 % correct	1998 Year 5 % correct
93	96

An example of one of the difficult mathematics items for Year 5 students in 1994 and 1998.

Example 2

Content area: *Measurement, and Estimation and Number Sense (solving problems)*

L8. Four children measure the width of a room by counting how many paces it took them to cross it. The chart shows their measurements.

Name	Number of Paces
Stephen	10
Erin	8
Anna	9
Jason	7

Who had the longest pace?

A. Stephen
B. Erin
C. Anna
D. Jason

1994 Year 5 % correct	1998 Year 5 % correct
21	22

An example of a mathematics item that exhibited no change for Year 5 students from 1994 to 1998

Example 3

Content area: *Whole Numbers (routine procedures)*

S2. Here is a number sentence.

2000 + ■ + 30 + 9 =2739

What number goes where the ■ is to make this sentence true?

1994 Year 5 % correct	1998 Year 5 % correct
49	49

Examples of mathematics items that exhibited a change for Year 5 students from 1994 to 1998

Example 4

Content area: *Data Representation, Analysis, and Probability (Knowing)*

L1. The graph shows 500 pine trees and 150 rimu Trees.

How many trees does each  represent

Answer:_____

1994 Year 5 % correct	1998 Year 5 % correct
45	55

Example 5

Content area: *Geometry (Knowing)*

L5. This picture shows a cube with one edge marked. How many edges does the cube have altogether?

A. 6
B. 8
C. 12
D. 24

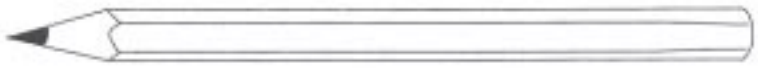
.

1994 Year 5 % correct	1998 Year 5 % correct
32	42

Example 6

Content area: *Measurement, and Estimation and Number (using complex procedures)*

K5. About how long is this picture of a pencil?



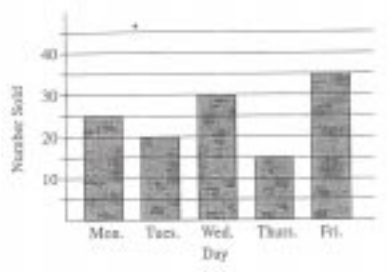
A. 5 cm
B. 10 cm
C. 20 cm
D. 30 cm

1994 Year 5 % correct	1998 Year 5 % correct
71	59

Example 7

Content area: *Data Representation, Analysis and Probability (solving problems)*

T1. The graph shows the number of cartons of milk sold each day of a week at a school.



How many cartons of milk did the school sell on Monday?

Answer: _____

1994 Year 5 % correct	1998 Year 5 % correct
77	85

How many cartons of milk did the school sell that week?
Show your working.

Answer:_____

1994 Year 5 % correct	1998 Year 5 % correct
40	44

Examples of mathematics items demonstrating gender differences for Year 5 students from 1994 to 1998

Example 8

Content area: *Fractions and Proportionality (using complex procedures)*

I5. Marcus uses 5 tomatoes to make half a litre of tomato sauce. How much sauce can he make from 15 tomatoes?

A. One and a half litres
B. Two litres
C. Two and a half litres
D. Three litres

	1994 Year 5 % correct	1998 Year 5 % correct
Girls	51	47
Boys	45	51
All **students**	48	49

Example 9

Content area: *Measurement, and Estimation and Number sense (solving problems)*

L6. The weight (mass) of a clothes peg is 9.2 grams. Which of these is the best estimate of the total weight of 1000 clothes pegs?

A. 900 g
B. 9 000 g
C. 90 000 g
D. 900 000 g

	1994 Year 5 % correct	1998 Year 5 % correct
Girls	42	56
Boys	42	47
All **students**	42	52

Example 10

Content area: *Whole Numbers (solving problems)*

L7. In which pair of numbers is the second number 100 more than the first number?

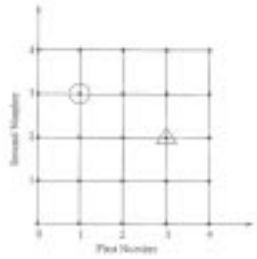
A. 199 and 209
B. 4236 and 4246
C. 9635 and 9735
D. 51 863 and 52 863

	1994 Year 5 % correct	1998 Year 5 % correct
Girls	29	27
Boys	36	47
All **students**	33	37

Example 11

Content area: *Geometry (solving problems)*

M4. On this grid, find the dot with the circle around it. We can describe where this is by saying it is at First Number 1, Second Number 3.



Now find the dot with the triangle around it. Describe where the dot is on the grid in the same way. Fill in the numbers we would use:

First Number_____ Second Number _____

	1994 Year 5 % correct	**1998 Year 5 % correct**
Girls	52	61
Boys	38	53
All **students**	45	57