



Third International Mathematics and Science Study – Repeat (TIMSS-R) Preliminary Achievement Results

Research Division

May 2000

It is now over a year since the Third International Mathematics and Science Study – Repeat (TIMSS-R) survey was administered in New Zealand. This report begins with a brief recap of the background to the study, followed by the preliminary achievement results for Year 5 and Year 9 students. In addition, a brief description of one of the technical aspects currently being undertaken at the international level – *scaling* – is also presented.

Key Points

- TIMSS-R is an international study of mathematics and science involving Year 9 students¹. New Zealand has also administered a national version of TIMSS-R at the middle primary level, involving mostly Year 5 students.
- The preliminary achievement results for each group of students are based on the performance on a subset of questions for each subject area, collectively known as the 'trend test items', that were common to both TIMSS (1994/95) and TIMSS-R (1998/99).
- They should only be viewed as an indication of the final results.
- The final results are not yet available, as the international researchers have not yet completed *scaling* the achievement data.
- Based on analyses using the trend test items for each subject area:
 - on average, the performance of Year 5 and Year 9 students in mathematics in 1998, was about the same as the performance of Year 5 and Year 9 students in 1994;
 - on average, the performance of Year 9 students in science in 1998 was about the same as the performance of Year 9 students in 1994;
 - on average, the performance of Year 5 students in science in 1998 was slightly higher compared to the performance of Year 5 students in 1994.
- The significance and size of any differences in achievement for particular subgroups between 1994 and 1998, as well as differences across subgroups will not be explored until the scaled scores become available.
- The international results for the lower secondary level are expected to be released by Boston College in the United States in early December 2000.

¹ TIMSS-R was administered in Southern Hemisphere countries in late 1998 and in Northern Hemisphere countries in early 1999.

TIMSS-R is a repeat of TIMSS, a study on mathematics and science. TIMSS-R is designed to measure trends in student performance as well as examining the contexts for learning mathematics and science.

Researchers in the Ministry of Education responsible for administering TIMSS-R in New Zealand work with a number of research organisations during different phases of the study.

In New Zealand, approximately 6900 Years 5 and 9 students took part in TIMSS-R.

Twenty-six of the 38 countries taking part in TIMSS-R also took part in TIMSS.

What is TIMSS-R?

TIMSS-R is a partial replication of TIMSS, a major international study on mathematics and science administered in about 40 countries, including New Zealand, during 1994–1995. TIMSS-R, administered in New Zealand in 1998, is not as extensive as TIMSS. At the international level, TIMSS-R is investigating the mathematics and science achievement of students only at the class level equivalent to form 3 (Year 9 students). New Zealand also chose to examine the mathematics and science achievement of students at the class level equivalent to standard 3 (mostly Year 5 students).²

By including both Year 5 and Year 9 students in TIMSS-R, New Zealand will be able to compare:

- Year 5 and Year 9 students with their counterparts who were assessed in TIMSS in 1994 (*historical* trend); and
- Year 9 students with their counterparts overseas who were assessed in TIMSS-R in 1998/99 (*international* comparison).

Who carried out TIMSS-R?

The Comparative Education Research Unit is responsible for carrying out the TIMSS-R activities in New Zealand. This unit is located within the Research Division of the Ministry of Education.

Boston College's School of Education manages the international coordination of the project, in the United States. The other organisations involved with the TIMSS-R activities are:

- Statistics Canada in Ottawa;
- the IEA Data Processing Centre in Hamburg (Germany); and
- Educational Testing Service in Princeton, New Jersey (United States).

Who took part in TIMSS-R in New Zealand?

In New Zealand, TIMSS-R has involved schools with:

Year 9 (form 3) students:

- 27 schools in the field trial;
- 152 schools in the main survey.

Year 5 (standard 3) students:

- 149 schools in the main survey.

Approximately 2400 Year 5 students and 4500 Year 9 participated in the study. The mathematics and science teachers of the students, as well as their principals also provided information on the context in which mathematics and science learning was taking place in their schools.

What countries or education systems are taking part in the international study?

Thirty-eight countries or education systems are taking part in TIMSS-R. These are:

Australia*	Hong Kong, SAR*	Lithuania*	Singapore*
Belgium (Flemish) *	Hungary*	Macedonia	Slovak Republic*
Bulgaria*	Indonesia	Malaysia	Slovenia*
Canada*	Iran*	Moldova	South Africa*
Chile	Israel*	Morocco	Thailand*
Chinese Taipei	Italy	The Netherlands*	Tunisia
Cyprus*	Japan*	New Zealand*	Turkey
Czech Republic*	Jordan	Philippines*	United States*
England*	Korea*	Romania*	
Finland	Latvia*	Russian Federation*	

* Countries that participated in TIMSS in 1994-1995.

² For schools that had moved to the new nomenclature the level defined was the class in which students had three further years of primary education before entering secondary school in 2002.

Three approaches will be taken to describe student performance in the TIMSS-R assessment.

How will student achievement results in TIMSS-R be reported?

There are three ways of reporting student achievement results:

1. Reporting on how students have responded to individual mathematics or science items – eg, the percentage of students who answered each individual item correctly. This approach was used in TIMSS, and examples have been used in this report.
2. Reporting an average score for sets of items, commonly referred to as the mean % correct score. This approach was used in TIMSS, in both international and national reports. It has been used to summarise the performance on the trend test items for each subject area in this report.
3. Reporting Item Response Theory (IRT) scaled scores, using information on the characteristics of both the test items and students taking the test. This approach was used to describe student performance in the international reports on TIMSS and will be used when reporting international and national TIMSS-R results. The *IRT* scaled scores are not yet available to present in this report. (See page 14 for more details).

How will historical trends be measured at the middle primary level?

The test used with the middle primary students was an exact replication of the one used in TIMSS. However, about two-thirds of the test questions were released into the public domain in 1997. Since no new test items were developed to replace the released items, two approaches will be taken to describe and compare student performance in 1994 and 1998:

- reporting *IRT* scaled scores, which uses all test items, for each subject area; and
- using the mean % correct approach to examine the performance of the two groups of students on the trend test items that were kept secured.

The performance of Year 5 students on the released items, many of which are of open-ended format, will also be scrutinised.

How will historical trends be measured at the lower secondary level?

Comparing Year 9 students with their TIMSS counterparts is more straightforward. As was the case with the middle primary level, about two-thirds of the TIMSS test questions were released into the public domain; one-third of the test questions were kept secured and were used for TIMSS-R. New test questions were developed to replace the released items. The psychometric scaling techniques employed at the international level enable the entire item pool to be used to measure trends between 1994 and 1998.

Why are the results presented in this report still preliminary?

As mentioned above, to be able to report **reliable** trend measures we will be using *IRT* scaled scores in the next reports. However, ETS in the United States have not yet completed the scaling of the 1998/1999 TIMSS-R achievement data. Therefore, only the mean % correct scores calculated across the pool of trend (or secured) items can be presented. This approach means there is still a certain level of uncertainty about the initial findings.

For example: while the mean % correct method provides a measure of performance across test items, it is not possible to describe the distribution of the scores for Year 5 and Year 9 students, since not all students responded to the same sets of test items. In contrast, *scaling* allows us to examine the distribution of scores.

For preliminary reporting, the mean % correct score across the set of trend test items has been calculated for each subject area.

There are some important technical factors to consider when looking at these preliminary results.

On each graph, error bars ($\approx \pm 2 \times$ standard error) or confidence intervals are shown for the means. If the error bars do not overlap then the means are likely to be (statistically) different from each other. However, any differences and the magnitude of the differences need to be examined with appropriate statistical testing.

Based on analyses using the trend test items...

... Years 5 and 9 students' performance in mathematics in 1998 was probably about the same as their counterparts in 1994.

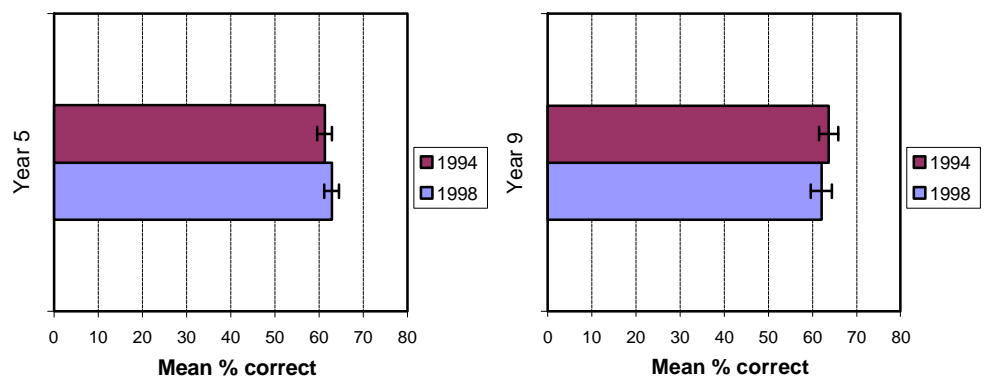
Some technical notes:

1. To have some indication of the level of uncertainty between the mean scores based on the samples of students and the actual means if every student in each population had answered all test items, standard errors were calculated for each of the means.
2. Both TIMSS and TIMSS-R used multistage cluster sampling – ie, schools were selected first and then a class or grouping was sampled. It is, therefore, inappropriate to use any assumptions of simple random sampling. A technique known as 'jackknife repeating replication' was used to calculate these standard errors.
3. It is important to remember that there is no "pass mark" (eg, 50%) in this assessment. Rather, the results are used to compare the performance of one group of students with another group (eg, girls' performance relative to that of boys; New Zealand students' performance relative to that of other English-speaking countries.)
4. For the purpose of this summary, trends in performance are examined within larger groupings – 'boys', 'girls', etc. When the *IRT* scaled scores become available, the performance of 'sub-groups' within each of these larger groupings will be examined, as well as the trends in performance across groups.

How does student performance in mathematics in 1998 compare with student performance in 1994?

As shown in Figure 1, the mean for the mathematics trend test items for Year 5 students in 1998 was slightly higher than the 1994 mean; the mean for Year 9 students in 1998 was a little lower than that achieved by their 1994 counterparts. Based on this finding, it is unlikely that the overall mean performance of both Year 5 and Year 9 students in 1998 will be significantly different from the performance of their respective 1994 counterparts.

Figure 1: Mean % correct for mathematics trend test items



Since the Year 5 cohort in standard 3 in 1994 were Year 9 students in form 3 in 1998, it seems likely, based on the first round of analyses, that this group in four years, had progressed to the level of their 1994 Year 9 counterparts. In TIMSS (1994/95) it was found that the 'apparent growth', estimated for New Zealand between middle primary and lower secondary level in mathematics, as measured by the change in New Zealand's standing relative to the international means, was considered 'high' when compared with other countries.

... Year 5 boys in 1998, appear to have performed at a slightly higher level in mathematics than their 1994 male counterparts.

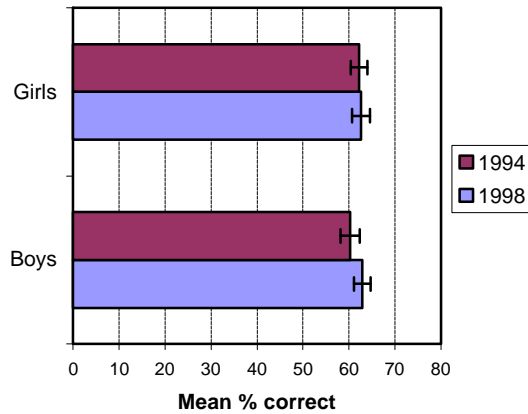
... Year 9 boys seem to have performed at a slightly lower level in mathematics in 1998 than Year 9 boys did in 1994.

... girls' mean performance in mathematics in both year groups in 1998 seems to be about the same as in 1994.

Gender

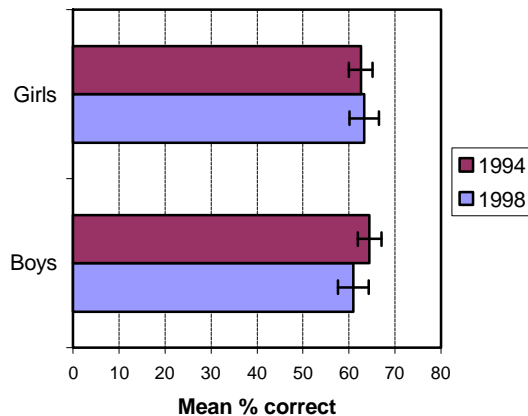
Year 5 girls' performance on the mathematics trend test items in 1998 was, on average, about the same as Year 5 girls in 1994. Year 5 boys appear to have, on average, fared better than their 1994 counterparts. However, the significance of this result for boys, if at all, will be determined once the *IRT* scaled scores become available.

Figure 2a: Year 5 students' mean % correct for the mathematics trend test items, by gender



On average, the performance of Year 9 girls on the trend test items in 1998 was not dissimilar to the performance of Year 9 girls' in 1994. However, Year 9 boys, on average, attained a lower mean score across the trend items in 1998 than their counterparts did in 1994. The significance of this difference in performance (ie, is this a statistically significant result? What is the magnitude of the difference?) will be explored further once the *IRT* scaled scores become available.

Figure 2b: Year 9 students mean % correct for mathematics trend test items, by gender



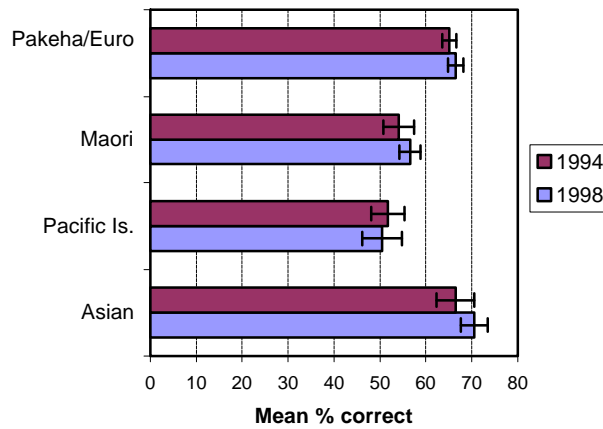
In TIMSS (1994/95) at the class level equivalent to standard 3, New Zealand was one of eight (out of 26) countries to observe higher mean mathematics scores for girls than for boys. It is possible that this trend has continued over the four years and hence the reason why boys' performance relative to that of girls is, on average, lower in 1998 than in 1994 where boys on average performed at a higher level compared to girls.

Ethnicity

Small increases in the mean scores for the mathematics trend test items from 1994 to 1998 were observed for Year 5 students in three of the four main ethnic groupings. In particular, the Maori and Asian groupings. The 1998 mean for

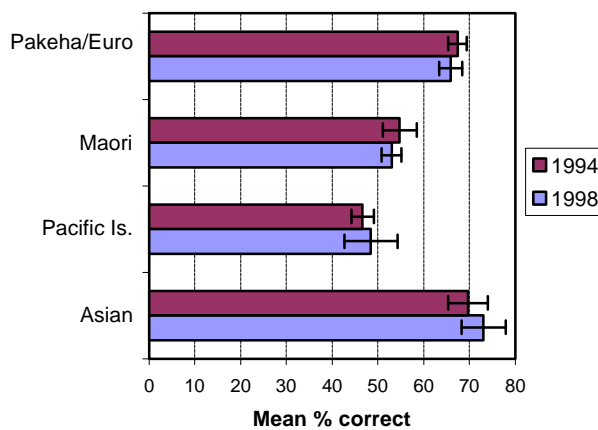
students in the Pacific Islands grouping was marginally lower than the 1994 mean. It is likely, however, that the overall 1998 means for the four groupings will not be markedly different from the 1994 means.

Figure 3a: Year 5 students' mean % correct for the mathematics trend test items, by ethnic grouping



At the lower secondary level, the mean scores on the trend test items for Year 9 students in the Asian and Pacific Islands groupings were higher in 1998 than in 1994; the mean scores for students in the Pakeha and Maori groupings were lower in 1998 than in 1994. However, based on this preliminary result it is unlikely that there will be any marked differences in their overall mathematics performance.

Figure 3b: Year 9 students mean % correct for the mathematics trend test items, by ethnic grouping



How does student performance in science in 1998 compare with student performance in 1994?

As shown in Figure 4, in 1998 Year 5 students, on average, attained higher scores on the science trend test items than Year 5 students did in 1994. The significance of this finding (eg, does this reflect a significant increase in mean science achievement?) will be explored further once the *IRT* scaled scores become available. At the lower secondary level, based on the performance on the trend test items, it is likely that the performance of Year 9 students in 1998 was not dissimilar to the performance of Year 9 students in 1994.

... Years 5 and 9 students in each of the four ethnic groupings in 1998 probably performed at about the same level in mathematics as their counterparts in 1994.

Based on analyses using the trend test items ...

... Year 5 students' performance in science appears to have been at a slightly higher level compared with Year 5 students' performance in 1994.

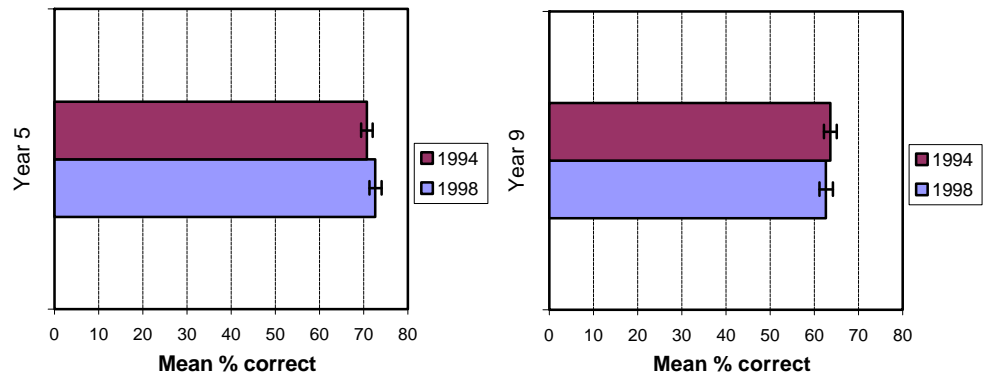
... Year 9 students' performance in science in 1998 seems to be at about the same level as their 1994 counterparts.

... Year 5 boys' performance in science in 1998 appears to be at a higher level than Year 5 boys' performance in 1994.

... Year 9 boys in 1998 appear to have performed at a lower level in science compared with Year 9 boys in 1994.

... Year 5 and Year 9 girls' performance in science in 1998 seems to be about the same as it was in 1994.

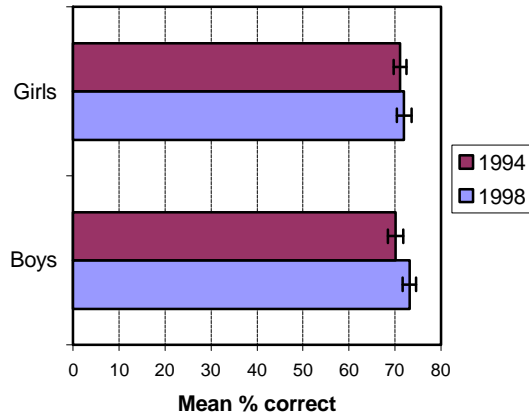
Figure 4: Mean % correct on science trend test items



Gender

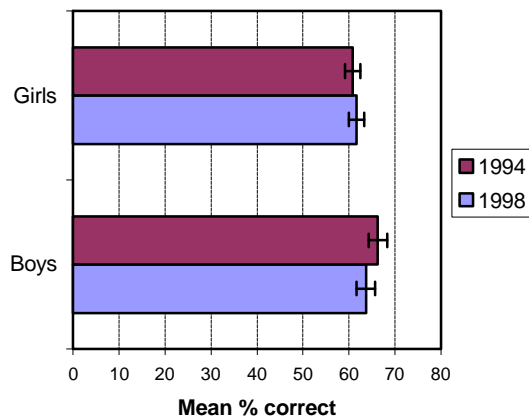
As was the case in mathematics, preliminary analyses show that on average both Year 5 girls and boys in 1998 performed at a slightly higher level on the science trend items than their respective counterparts in 1994. However, it is worth noting the apparent increase in means for Year 5 boys between 1994 and 1998.

Figure 5a: Year 5 students mean % correct for the science trend test items, by gender



In contrast to the apparent increase in the means observed for boys at the middle primary level, the converse was observed for Year 9 boys. That is, Year 9 boys, on average, attained slightly lower scores on the science trend test items than their Year 9 counterparts in 1994.

Figure 5b: Year 9 students mean % correct for the science trend test items, by gender



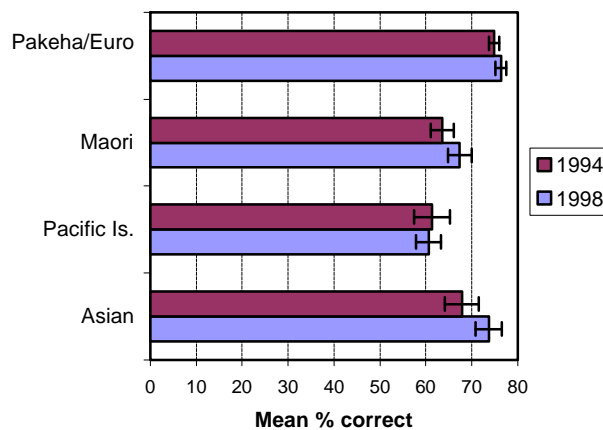
By way of contrast, Year 9 girls' performance in 1998, on average, appears to be not dissimilar to that of Year 9 girls in 1994.

In TIMSS (1994/95) at the form 3 level, New Zealand recorded the fourth largest (statistically significant) gender difference, which favoured boys, out of 42 countries. Preliminary results presented here would suggest that by 1998, the difference between boys and girls had decreased markedly, and it is likely that there will be no significant difference between their overall performance.

Ethnicity

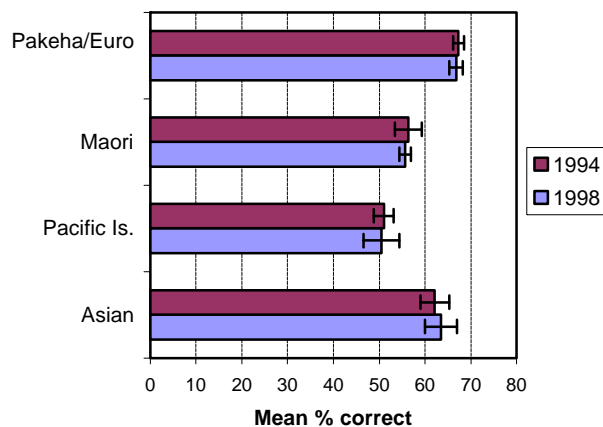
As was the case for mathematics at the middle primary level, the mean scores on the science trend test items for Year 5 students in three of the four ethnic groupings were higher in 1998 than in 1994, although the mean for Pakeha was only marginally so. The 1998 mean for the Pacific Islands grouping seems to be about the same as the 1994 mean. Once the *IRT* scaled scores become available, a closer examination of the apparent positive change particularly for Maori and Asian will be undertaken. It seems unlikely, however, that the overall means for the Pakeha and Pacific Islands groupings in 1998 will be notably different from the 1994 means for these two groupings.

Figure 6a: Year 5 students mean % correct for science trend test items, by ethnic grouping



In contrast to the more positive findings at the middle primary level, there was very little change — improvement or deterioration — for Year 9 students in each of the four main groupings in science over the four years.

Figure 6b: Year 9 students mean % correct for the science trend test items, by ethnic grouping



... Year 5 students in two of the four main ethnic groupings in 1998 appear to have performed at slightly higher levels than their 1994 counterparts.

... Year 9 students in each of the four main ethnic groupings seem to have performed at about the same level in science as their 1994 counterparts.


Examples of some of the test items used at the middle primary and lower secondary levels

Mathematics items

These two mathematics questions are examples of items that showed an increase in the proportions of Year 5 students answering correctly.

Example 1: answered by Year 5 students

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

Copyright © by IEA, Amsterdam.

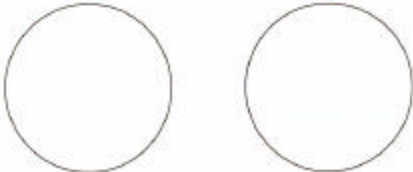
1994 Yr 5 % correct	1998 Yr 5 % correct
32	42

Example 2: answered by Year 5 students

V1. Sam said that $\frac{1}{3}$ of a pie is less than $\frac{1}{4}$ of the same pie.

Is Sam correct? _____

Use the circles below to show why this is so.



Shade in $\frac{1}{3}$ of this circle

Shade in $\frac{1}{4}$ of this circle

Copyright © by IEA, Amsterdam

Marks for correctness	1994 Yr 5% correct	1998 Yr 5 % correct
At least one mark	47	54
Two marks	16	23

These two mathematics questions are examples of secured items presented to Year 9 students in both 1994 and 1998.

Example 3: answered by Year 9 students

B8. If there are 300 calories in 100 g of a certain food, how many calories are there in a 30 g portion of food?

A. 90
B. 100
C. 900
D. 1000
E. 9000

Copyright © by IEA, Amsterdam.

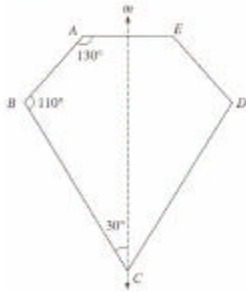
1994 Yr 9 % correct	1998 Yr 9 % correct
67	69

Example 4: answered by Year 9 students

D7. The line m is a line of symmetry for figure $ABCDE$.

The measure of angle BCD is

A. 30
B. 50
C. 60
D. 70
E. 110



Copyright © by IEA, Amsterdam

1994 Yr 9 % correct	1998 Yr 9 % correct
59	54

These two science questions are examples of items that showed an increase in the proportions of Year 5 students answering correctly.

Science items

Example 1: answered by Year 5 students

- Q9. Ken put a thermometer in a glass filled with hot water. Why does the liquid inside the thermometer rise?
- A. Gravity pushes it up.
 - B. Air bubbles are released.
 - C. Heat from the water makes it expand.
 - D. Air pressure above the water pulls it up.

Copyright © by IEA, Amsterdam.

1994 Yr 5 % correct	1998 Yr 5 % correct
59	63

Example 2: answered by Year 5 students

- Y1. The Sun is bigger than the Moon, but they appear to be about the same size when you look at them from the Earth. Why is this?

Copyright © by IEA, Amsterdam.

1994 Yr 5 % correct	1998 Yr 5 % correct
34	41

These two science questions are examples of secured items presented to Year 9 students in both 1994 and 1998.

Example 3: answered by Year 9 students

- B2. Most of the chemical energy released when petrol burns in a car engine is not used to move the car, but changed into
- A. electricity
 - B. heat
 - C. magnetism
 - D. sound

Copyright © by IEA, Amsterdam.

1994 Yr 9 % correct	1998 Yr 9 % correct
58	58

Example 4: answered by Year 9 students

- F6. Paint applied to an iron surface prevents the iron rusting. Which ONE of the following provides the best reason.
- A. It prevents nitrogen from coming in contact with the iron.
 - B. It reacts chemically with the iron.
 - C. It prevents carbon dioxide from coming in contact with the iron.
 - D. It makes the surface of the iron smoother.
 - E. It prevents oxygen and moisture from coming in contact with the iron.

Copyright © by IEA, Amsterdam.

1994 Yr 9 % correct	1998 Yr 9 % correct
70	66

These four items illustrate the change in the proportions of boys and girls answering correctly in 1994 and 1998.

Examples of items that exhibited change in the proportion of girls and boys answering correctly

Example 1: answered by Year 5 students

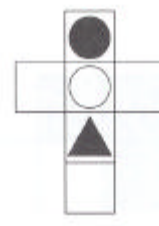
- I5. Marcus uses 5 tomatoes to make half a litre of tomato sauce. How much sauce can he make from 15 tomatoes?
- A. One and a half litres
 - B. Two litres
 - C. Two and a half litres
 - D. Three litres

Copyright © by IEA, Amsterdam.

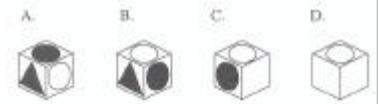
	1994 Yr 5 % correct	1998 Yr 5 % correct
Girls	51	47
Boys	45	51

Example 2: answered by Year 9 students

B11.



Which of these cubes could be made by folding the figure above?

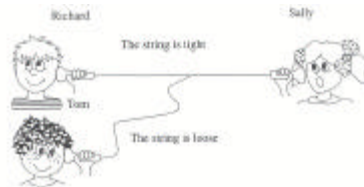


Copyright © by IEA, Amsterdam.

	1994 Yr 9 % correct	1998 Yr 9 % correct
Girls	70	71
Boys	74	69

Example 3: answered by Year 5 students

- R5. The picture shows Richard and his friends playing with a string-telephone. Sally is speaking. Richard and Tom are trying to listen. Which of them can hear her speak?



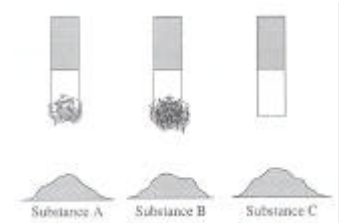
- A. Both of them can hear equally clearly.
- B. Neither of them can hear.
- C. Only Tom can hear clearly.
- D. Only Richard can hear clearly.
- E. Both of them hear equally faintly.

Copyright © by IEA, Amsterdam.

	1994 Yr 5 % correct	1998 Yr 5 % correct
Girls	55	43
Boys	40	46

Example 4: answered by Year 9 students

- D2. Each of the three magnets has been dipped into the substance below it. Which of the substances could be coffee?



- A. A only
- B. B only
- C. C only
- D. A and B only.

Copyright © by IEA, Amsterdam.

	1994 Yr 9 % correct	1998 Yr 9 % correct
Girls	70	71
Boys	79	75

A comparison of demographic characteristics between 1994 and 1998 shows ...

... there were about the same proportion of boys and girls in both studies.

... small changes in the ethnic composition of the student groups between 1994 and 1998.

How do the demographic characteristics of the TIMSS-R students in 1998 compare with the TIMSS students in 1994?

In addition to examining the trends in student achievement, trends in the contextual information gathered from students, teachers, and principals will also be examined. For this preliminary report, only the demographic characteristics of Year 5 and Year 9 students (in standards 3 and form 3 respectively) in 1994 and 1998 are presented³.

Gender

Table 1 presents a breakdown of the student groups in the two studies by gender. The proportion of girls at each educational level was about the same in each study. Similarly, the proportions of boys at each level were also about the same.

Table 1: *The proportions of girls and boys in TIMSS and TIMSS-R*

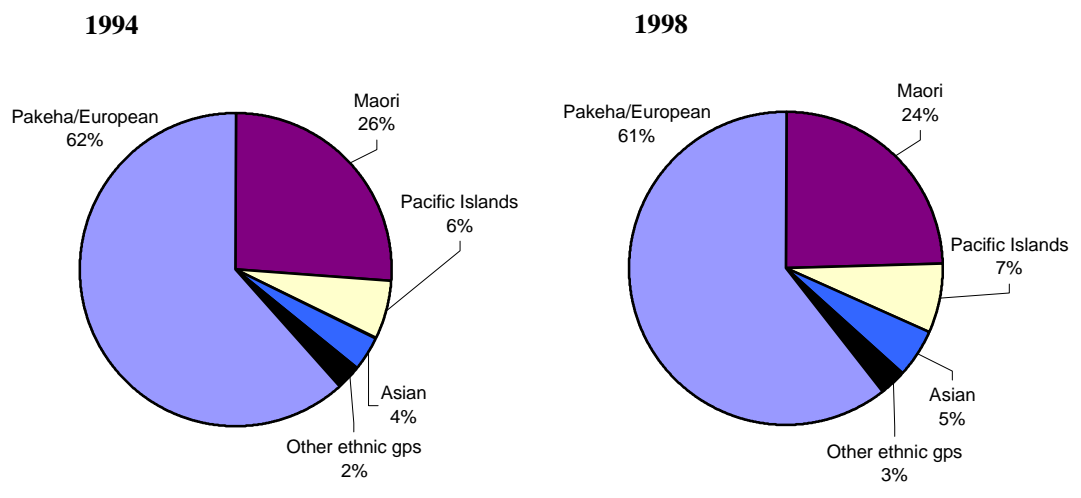
Students	1994		1998	
	% Girls	% Boys	% Girls	% Boys
Year 5	51	49	50	50
Year 9	48	52	49	51

Ethnic identity

Students were asked with which ethnic group or groups they identified.⁴ Their responses were organised into five groupings. These were *Pakeha/European*, *Maori*, *Pacific Islands*, *Asian*, and *Other Ethnic Groups*. Figures 7a and 7b show the proportions of Year 5 and Year 9 students in each of these groupings.

The proportions of Year 5 students in each of the groupings were essentially the same in both 1998 and in 1994, with only some very small changes for the *Pakeha/European* and *Maori* groupings.

Figure 7a: *Proportion of Year 5 students in each main ethnic grouping*

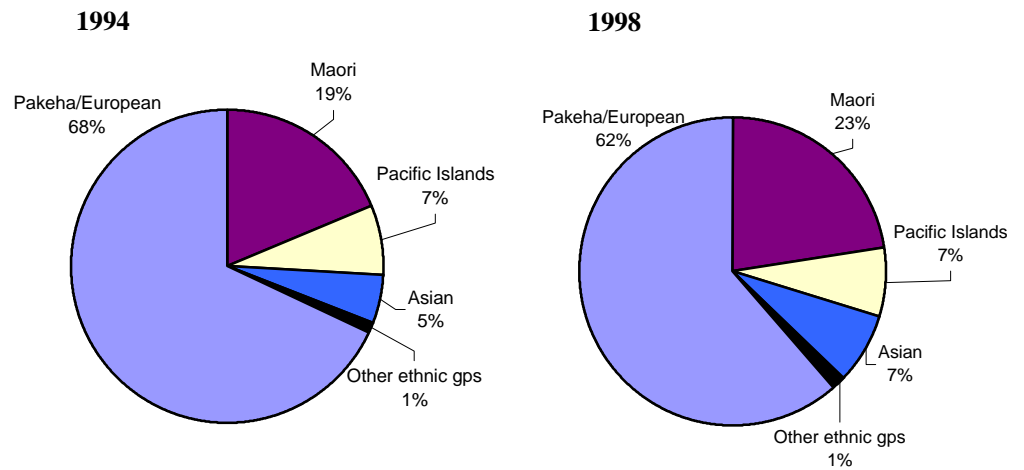


³ The responses from students were adjusted (weighted) to represent the student populations from which they were drawn.

⁴ The New Zealand Department of Statistics Standard Classification of Ethnicity (1992, 1996), a hierarchical classification, was used to group student responses.

At the lower secondary level, the most notable increases in proportions between 1994 and 1998 were observed for Maori and Asian ($\approx 3\%$ difference). By way of contrast, the proportion of students in the Pakeha/European grouping in 1998 was 6% lower than in 1994.

Figure 7b: *Proportion of Year 9 students in each main ethnic grouping*



... only very small changes in the proportions of students reporting that they were born overseas and spoke a language other than English in the home.

Country of origin

Table 2 shows that in 1998 slightly higher proportions of students indicated that they were born in a country other than New Zealand compared with students in 1994.

Table 2: *Proportion of students born in a country other than New Zealand*

Students	1994 (%)	1998 (%)
Year 5	10	12
Year 9	12	14

Language of the home

Slightly higher proportions of students in 1998 than in 1994 reported that they spoke a language other than English in the home at least some of the time. This was more so at the middle primary level than at the lower secondary level.

Table 3: *Proportion of students speaking a language other than English in the home*

Students	1994 (%)	1998 (%)
Year 5	13	15
Year 9	9	10

Scaling enables analyses to be carried out when students have been tested on different subsets of items from a larger item pool.

There are four main steps taken to generate the IRT scaled scores.

What has actually happened to the data at the international level?

Once every country had completed the scoring of open-ended questions, entered all questionnaire and test information into the databases, the datafiles were sent to the Data Processing Centre (DPC) in Germany for ‘cleaning’. The DPC were responsible for checking all countries’ datafiles, before scaling could be started.

ETS in the United States are now in the process of *scaling* the 1998/99 achievement data.

Why do we scale?

Since each student responded to only a subset of items from the entire mathematics test item pool and from the entire science item pool, scaling makes it possible to summarise their performance for each subject area on a common scale. The technique takes into account the differences in the difficulty level of the items answered by the students. The common scales make it possible to examine the relationships between students’ characteristics and their overall performance in mathematics and science.

It is also important to remember that TIMSS-R is concerned with estimating the performance of the Years 5 and 9 student populations (and subgroups within these populations) and is not interested in the performance of individual students.

What is actually involved in the scaling process?

The scaling process outlined here uses data from all 38 TIMSS-R countries. The process is used to find out if any items need to be dropped from further analyses as well as to generate scores.

Essentially, the steps are:

1. *Differential Item Functioning* analysis — this procedure determines how comparable the test items are across countries. It reveals whether any of the items work sufficiently differently across the various participating countries to warrant elimination. This is a very important step in assessment projects where you are comparing the performance of students in a range of countries. This analysis was carried out after the field trial and will also be carried out again prior to steps 2 to 4 being carried out.
2. *Scaling* — each item in the assessment has a number of ‘attributes’ that can be described both in words and mathematically. For example, ‘item difficulty’ is an attribute common to multiple choice, short answer and extended response items. Using Item Response Theory (IRT), the ‘attributes’ of an item are calculated and then combined mathematically; this procedure enables us to determine each item’s effectiveness to distinguish between levels of student proficiency in the subject area that item is assessing.
3. *Conditioning* — this is the step from which we can estimate the performance of the student population under scrutiny. It combines student achievement and background data, and the mathematical ‘attributes’ for each item calculated in step 2. The outcome from this step is a set of scaled scores.
4. *Transforming the scaled scores on to a reporting scale* — for reporting purposes the scaled scores that are estimated from step 3 will be standardised to have a mean of 500 and standard deviation of 100.

The TIMSS achievement data from 1994 has been rescaled so that it is on reporting scales comparable to the 1998 TIMSS-R achievement data.

Re-scaling TIMSS 1994/95

In addition to the need for scaling the 1998/99 achievement data there has been an added complication: the need to rescale the TIMSS achievement data. This was necessary because each TIMSS student population under scrutiny extended over two adjacent class levels — equivalent to standards 2 and 3 and forms 2 and 3. The scaling process for TIMSS involved students from each class level. The TIMSS-R student populations were defined over just one year of schooling. This meant any trends on student performance needed to be measured on a common scale.

ETS have recently completed re-scaling the international 1994/95 achievement data for middle primary and lower secondary using just the achievement information for the class level equivalent to standard 3 (grade 4) and form 3 (grade 8) respectively.

What next?

Boston College, in the United States, are planning to release findings from the international study on 5 December 2000 (Eastern Standard Time); in New Zealand this will be 6 December 2000.

We are planning to release final results for both Year 5 and Year 9 students to coincide with the release of the international reports.

Further information

For additional details about TIMSS or TIMSS-R go to www.timss.org or www.timss.bc.edu or

Enquiries about this project may be directed to:

Megan Chamberlain
National Research Coordinator, TIMSS-R
Research Division
Ministry of Education
Box 1666
Wellington
e-mail: megan.chamberlain@minedu.govt.nz
phone: 0-4-471 6107 or
fax: 0-4-471 4412.