MINISTRY OF EDUCATION NEW ZEALAND
*Te Tāhuhu o te Mātauranga Aotearoa*

# science

## trends in year 5 science achievement 1994 to 2006

New Zealand results from three cycles of the Trends in International Mathematics and Science Study (TIMSS)

Robyn Caygill

IEA TIMSS 2006/07

New Zealand Government

# Overview of TIMSS

## What is TIMSS?

The Trends in International Mathematics and Science Study (TIMSS) measures trends in mathematics and science achievement at the fourth and eighth grades (Years 5 and 9) as well as monitoring curricular implementation and identifying the most promising instructional practices from around the world.

Conducted on a regular 4-year cycle, TIMSS has assessed mathematics and science in 1994/95[1], 1998/99, 2002/03, and 2006/07 with planning underway for 2010/11.

## What does TIMSS consist of?

TIMSS consists of assessments of students' achievements in mathematics and science along with questionnaires for students, teachers, and principals to gather background information. The background information provides a context within which the achievement can be examined.

The TIMSS assessments are organised around two dimensions: a content dimension specifying the domains or subject matter to be assessed within mathematics and science; and a cognitive dimension specifying the domains or thinking processes to be assessed. These domains are published in the *TIMSS 2007 assessment frameworks* (Mullis, Martin, Ruddock, O'Sullivan, Arora, and Erberber, 2005). To guide questionnaire development, the contextual factors associated with students' learning in mathematics and science are also included in the frameworks.

## How was TIMSS developed?

The TIMSS tests were developed cooperatively with representatives from those participating countries that have been involved throughout the entire process. Questions were field-tested with a representative sample of students in these countries and the results generated were used to select and refine the questions for the final test. Questions for the background questionnaires underwent a similar process.

## Who participated?

In TIMSS 2006/07, approximately 425,000 students in 59 countries from all around the world took part. Participants included 183,150 students from 37 countries and 7 benchmarking participants at the middle primary level, and 241,613 students from 50 countries and 7 benchmarking participants at the lower secondary level.[2] In this cycle of TIMSS, only Year 5 students from New Zealand participated.

## Who administered TIMSS?

A consortium was responsible for managing the international activities required for the project. This consortium comprised: the International Study Centre, Lynch School of Education at Boston College, (Massachusetts) United States; the IEA Secretariat in Amsterdam, the Netherlands; the IEA's Data Processing Centre in Hamburg, Germany; Statistics Canada in Ottawa, Canada; and the Educational Testing Service (ETS) in Princeton, New Jersey in the United States. In New Zealand the Comparative Education Research Unit in the Ministry of Education was responsible for carrying out TIMSS.

## What procedures were used to ensure the quality of the data?

TIMSS procedures are designed to ensure the reliability, validity, and comparability of the data through careful planning and documentation, cooperation among participating countries, standardised procedures, and attention to quality control throughout. Procedures included verification of translations and layout of booklets and questionnaires, monitoring of sampling activities, international and national quality control observers during test administration, checking of data, detailed manuals covering procedures, and rigorous training for all involved. Members of the consortium ensured procedures were adhered to by all participating countries.

## Why participate in TIMSS?

Although it is often assumed that the international studies are only useful for international benchmarking purposes, the real value of TIMSS lies in its ability to provide a rich picture of mathematics and science achievement within New Zealand and over time.

TIMSS (along with other international assessment studies) can provide information about the performance of the New Zealand education system at the national level within a global context. The information from studies such as TIMSS is used in the development and review of policy frameworks and also to inform and improve teaching practice. Developments arising out of previous cycles of TIMSS include resource materials for schools and teachers along with teacher in-service training programmes.

1 Note that this cycle of the study is called TIMSS 1995 internationally as most countries participated in 1995. However southern hemisphere countries conducted the assessment towards the end of 1994 so in New Zealand reports the study is referred to as TIMSS 1994/95. Similarly for the subsequent cycles, the two years in which administrations occurred in participating countries are indicated.

2 Mongolia does not appear in any international comparisons because they were unable to meet sampling criteria. Selected results for Mongolia appear in Appendix E of Martin, Mullis, and Foy (2008). Throughout the report 36, rather than 37, countries are discussed at the middle primary level.

# Table of Contents

# Acknowledgements

TIMSS is a collaborative effort internationally and nationally. There are many people that have been involved and it is not possible to thank all of them individually here. However, there are a few key people that I would like to acknowledge.

This study was made possible by the cooperation of the schools, teachers, and students that participated. Thanks to these participants, large amounts of data have been collected about the teaching and learning of New Zealand Year 5 students.

Thanks also to the many members of the Research Division, and also to those from the wider ministry, that have contributed to the successful collection of data for TIMSS, particularly Kate Lang. Special thanks are also extended to Megan Chamberlain and Ian Schagen for their help with the technical aspects of analysing the data.

Finally thanks to those that contributed to this publication by reviewing, editing and publishing this report. In particular, thanks to Sarah Kirkham for her support in getting this report written.

Robyn Caygill
TIMSS National Research Coordinator

# Key findings

## Achievement

- The mean science achievement of New Zealand Year 5 students was about the same in 2006 as in 1994. Although results from 1994, 1998, and 2002, showed a steady increase, this trend did not continue in 2006 when the results returned to the 1994 levels.
- New Zealand Year 5 mean science achievement was significantly[3] higher than 13 of the 36 countries that participated at the middle primary level.
- A comparison with the other countries that have taken part in TIMSS across all three of the cycles shows that the mean science achievement of New Zealand Year 5 students has moved little in relation to these countries.
- The range of New Zealand Year 5 science achievement was narrower in 2006 than in 1994, with fewer students demonstrating very high or very low achievement.
- Year 5 students demonstrated a relative strength in *earth science* questions compared to *life* and *physical* science. Students also performed relatively better on questions that involved *demonstrating* knowledge compared to *applying* knowledge or *reasoning*.

## Background context

- In 2006, teachers reported significantly fewer hours teaching science to New Zealand Year 5 students, on average, compared with 2002. The number of hours reduced from 66 per year in 2002 to 45 in 2006.
- There was no significant difference in mean science achievement between New Zealand Year 5 boys and girls.
- Both high and low performers were found in all ethnic groupings. Päkehä/European and Asian students had, on average, significantly higher mean science achievement than their Mäori and Pasifika counterparts. There was no difference in the average performance of Päkehä/European and Asian students. Mäori students had significantly higher mean science achievement than Pasifika students.
- Science achievement was higher, on average, among students who regularly spoke English at home. Students who were born in New Zealand had higher science achievement, on average, than those who were not.
- Students from higher socio-economic backgrounds tended to have higher mean science achievement than those from lower socio-economic backgrounds as evidenced by the proxy measures *books in the home*, *items in the home*, *household size* and *mobility*. In addition, the decile of the school they attended, indicative of the level of economic disadvantage in the community in which they live, was positively related to science achievement.
- Year 5 students who reported a small or moderate amount of time in out-of-school leisure activities generally had higher achievement than those who either reported no time or reported many hours on the activity.

## Student attitudes

- New Zealand Year 5 students generally expressed positive attitudes towards science. Eight out of every ten students indicated that they would like to do more science in school. Those students who reported positive attitudes towards science or were confident in their own science abilities had higher achievement than those who were less positive or confident.
- Boys and girls expressed similar attitudes to science, both in terms of enjoyment and motivation, and of self-confidence.
- More Päkehä/European and students in the Other ethnic grouping reported high self-confidence in science compared with Asian, Mäori, and Pasifika students. Proportionally more students in the Other ethnic grouping reported positive attitudes towards science compared with Päkehä/European, Asian, Mäori, and Pasifika students.

3 The term 'significantly' is used throughout this report to refer to statistical significance.

# Introduction

This report examines the science results for New Zealand Year 5 students from TIMSS 2006/07.[4] Along with the report on New Zealand's results for mathematics (Caygill & Kirkham, 2008), this report forms the beginning of a series of reports around New Zealand's participation in TIMSS 2006/07. International findings for science for TIMSS 2006/07 have been published by the IEA (Martin, Mullis & Foy, 2008). A separate international report on mathematics was also published at this time (Mullis, Martin & Foy, 2008).

This report begins by examining trends in New Zealand science achievement at the Year 5 level from 1994 to 2006. It then looks at New Zealand's science achievement in relation to other countries that participated in the study. An examination of the TIMSS assessment questions in relation to New Zealand's science curriculum is presented along with analyses of achievement by sub-groupings (such as gender and ethnicity) and background factors. Lastly, a statistical model that attempts to explain variations among students, classes, and schools, using the background information discussed in this report is also described.

## Assessment of science in TIMSS

The TIMSS assessment has two main dimensions: a dimension that describes the content or subject matter to be assessed; and a dimension that describes the cognitive processes used to answer the questions. The three content dimensions for science are: life science, physical science, and earth science. The detail about the topic areas covered in these domains at each grade or year level assessed and a set of assessment objectives for each topic area are presented in the *TIMSS 2007 assessment frameworks* (Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberber, 2005). Briefly, each of the content areas is described in the frameworks as follows.

*"Life science includes understandings of the characteristics and life processes of living things, the relationships between them, and their interaction with the environment."* (p. 43).

*"Physical science includes concepts related to matter and energy, and covers topics in the areas of both chemistry and physics. Since students in fourth grade have only a beginning knowledge of chemistry, the framework places more emphasis on physics concepts."* (p. 47).

*"Earth science is concerned with the study of Earth and its place in the solar system."* (p. 50).

In order to answer questions in the TIMSS test correctly, as well as being familiar with the science content, students need to draw on a range of cognitive skills. Also, in their lives outside and beyond school, students will need to do more than just accurately recall a range of science facts. This is acknowledged in the framework with three aspects to the cognitive dimension entitled knowing, applying, and reasoning. Briefly, each cognitive dimension is described in the framework as follows.

*"The first domain,* Knowing, *covers facts, procedures, and concepts students need to know, while the second domain,* Applying*, focuses on the ability of the student to apply knowledge and conceptual understanding in a problem situation. The third domain,* Reasoning, *goes beyond the solution of routine problems to encompass unfamiliar situations, complex contexts, and multi-step problems."* (p. 68).

4 Internationally this cycle of the study is called TIMSS 2007. As southern hemisphere countries conducted the study first, towards the end of 2006, while northern hemisphere countries conducted the study in 2007, it is referred to as TIMSS 2006/07 throughout this report.

## Data collection

Each student was assessed in two timed sessions of 36 minutes, and answered a combination of mathematics and science questions. The assessment was a pencil-and-paper test containing both multiple-choice and constructed-response questions. Following this, students were given a questionnaire containing questions about themselves, their opinions about science and mathematics, their computer use and time spent on homework. Principals and teachers were also given questionnaires in order to gain further information about the context in which the science teaching and learning take place. In New Zealand, the assessments and questionnaires were conducted in English.[5]

## International participants in TIMSS

The number of participants in TIMSS at the Year 5 or grade 4 level has steadily increased since 1994, when 26 countries took part. In 2002, 25 countries and 3 benchmarking participants took part. Benchmarking participants are usually states or parts of countries and are not included in international averages. In 2006, the number of education systems participating at the middle primary level had risen to 37 countries and 7 benchmarking participants.

## Technical information

A lot of information is gathered during the TIMSS administration and a number of techniques are applied when collecting and analysing the data. The *TIMSS 2007 technical report* (Olson, Martin, & Mullis, (Eds.), 2008) contains a detailed account of the procedures for scoring, translation of materials, sampling, survey operations, quality assurance, sampling weights, item analysis, scaling, and reporting. In addition, the *TIMSS 2007 user guide for the international database* (to be published in early 2009) contains information on how to analyse the data. Brief details of the technical information are given in the definitions and technical notes at the end of this report.

## TIMSS encyclopaedia

In order to provide a context in which the TIMSS results can be examined, TIMSS also publishes the *TIMSS 2007 encyclopedia: a guide to mathematics and science education around the world* (Mullis, Martin, Olson, Berger, Milne, & Stanco (Eds.), 2008). This encyclopaedia contains short reports from each country describing mathematics and science education policies and practices in that country.

5 In 2002, tests and questionnaires were also translated into te reo Mäori, but in order to make comparisons between each of the cycles, these students were excluded from analyses presented in this report.

# Trends in New Zealand science achievement 1994 to 2006

## Trends in means and ranges since 1994

New Zealand has participated in TIMSS since its inception in 1994. In 1998, although no assessment was offered internationally at the middle primary level, New Zealand opted to repeat the 1994 assessment. Therefore, we now have information from four different assessments of science achievement. Figure 1 presents the distributions of science achievement of New Zealand Year 5 students over the four cycles of TIMSS.

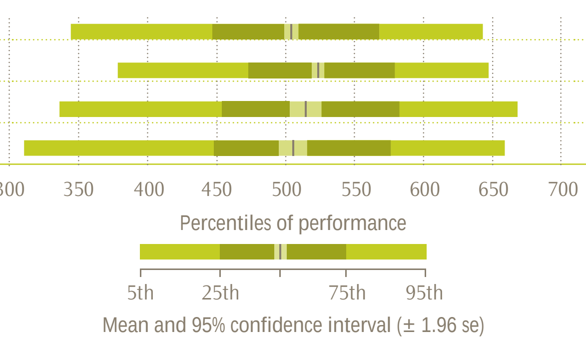
The results from an examination of science achievement since 1994 (see Figure 1) show that mean science achievement in 2006 is about the same as 1994, the first cycle of TIMSS. However, the 2006 mean is significantly lower (18 scale score points) than that of 2002.[6] Although the mean score for 2006 is numerically lower than 1998, the difference between 1998 and 2006 is not significant. The overall picture of trends over time when examining the mean science achievement of Year 5 students shows a steady increase from 1994 to 2002; however this pattern was not maintained in 2006 when the results returned to 1994 levels.

It is also useful to look at the range of achievement as represented by the outer limits of achievement. The lowest outer limit presented in Figure 1 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. In addition, the 25th and 75th percentiles are also presented in Figure 1, along with the inter-quartile range.

As shown in Figure 1, the range of achievement was narrower in 2006 than in both 1998 and 1994, but not as narrow as in 2002. A positive aspect of this change is that fewer students are demonstrating very low achievement, but, in addition, a smaller proportion of New Zealand students are gaining very high scores.

**Figure 1 Distribution of New Zealand Year 5 science achievement in TIMSS from 1994 to 2006**

| Year | Mean science score | Distribution of science achievement | Range from 5th to 95th percentile | Inter-quartile range from 25th to 75th percentile |
|---|---|---|---|---|
| 2006 | 504 (2.6) | | 299 | 121 |
| 2002 | 523 (2.3) | | 269 | 106 |
| 1998 | 514 (5.9) | | 332 | 129 |
| 1994 | 505 (5.3) | | 349 | 128 |




*Note: For trend purposes, only students tested in English are included in the results for 2002. Standard errors are presented in parentheses.*

## Trends in benchmarks for science

In order to describe more fully what achievement on the science scale means, the TIMSS international researchers have developed benchmarks. These benchmarks link student performance on the TIMSS science scale to performance on science questions and describe what students can typically do at set points on the science achievement scale. The international science benchmarks are four points on the science scale; the advanced benchmark (625), the high benchmark (550), the intermediate benchmark (475), and the low benchmark (400). The performance of students reaching each benchmark is described in relation to the types of questions they answered correctly. Table 1 presents the descriptions of the international benchmarks of science achievement.

**Table 1 TIMSS 2006/07 international benchmarks of science achievement**

| Advanced international benchmark – 625 |
|---|
| *Students can apply knowledge and understanding of scientific processes and relationships in beginning scientific inquiry.* Students communicate their understanding of characteristics and life processes of organisms as well as of factors relating to human health. They demonstrate understanding of relationships among various physical properties of common materials and have some practical knowledge of electricity. Students demonstrate some understanding of the solar system and Earth's physical features and processes. They show a developing ability to interpret the results of investigations and draw conclusions as well as a beginning ability to evaluate and support an argument. |
| **High international benchmark – 550** |
| *Students can apply knowledge and understanding to explain everyday phenomena.* Students demonstrate some understanding of plant and animal structure, life processes, and the environment and some knowledge of properties of matter and physical phenomena. They show some knowledge of the solar system, and of Earth's structure, processes, and resources. Students demonstrate beginning scientific inquiry knowledge and skills, and provide brief descriptive responses combining knowledge of science concepts with information from everyday experience of physical and life processes. |
| **Intermediate international benchmark – 475** |
| *Students can apply basic knowledge and understanding to practical situations in the sciences.* Students recognize some basic information related to characteristics of living things and their interaction with the environment, and show some understanding of human biology and health. They also show some understanding of familiar physical phenomena. Students know some basic facts about the solar system and have a developing understanding of Earth's resources. They demonstrate some ability to interpret information in pictorial diagrams and apply factual knowledge to practical situations. |
| **Low international benchmark – 400** |
| *Students have some elementary knowledge of life science and physical science.* Students can demonstrate knowledge of some simple facts related to human health and the behavioural and physical characteristics of animals. They recognize some properties of matter, and demonstrate a beginning understanding of forces. Students interpret labelled pictures and simple diagrams, complete simple tables, and provide short written responses to questions requiring factual information. |

*Source: Exhibit 2.1 from Martin, Mullis, and Foy, 2008.*

Table 2 presents the proportions of New Zealand Year 5 students that reached each of the benchmarks in each cycle from 1994 to 2006. Note that the proportion shown for the low benchmark also includes students who performed at the advanced, high, and intermediate benchmarks. This is because, by definition, students who could do the more complex questions associated with, for example, the high benchmark, would also be able to complete the easier questions associated with the intermediate and low benchmarks.

Eight percent of students reached the advanced benchmark in 2006, which was significantly fewer than in 1998 and 1994. While the proportion of students reaching the advanced benchmark peaked in 1998 (12%), the proportion of students reaching the high, intermediate and low benchmarks peaked in 2002 (39%, 74%, and 92% respectively). Significantly fewer students reached the high, medium and low benchmarks in 2006 compared with 2002.

There was also a group of Year 5 students in each cycle who did not reach the low benchmark. In terms of the benchmark definitions, these were students who did not demonstrate some elementary knowledge of life science and physical science. This group was proportionally largest in 1994 (15%) and smallest in 2002 (8%).

6 As mentioned in the introduction, only those students tested in English are included in trend comparisons.

**Table 2** **Trends in proportions of Year 5 students at each benchmark from 1994 to 2006**

| | Percentage of Year 5 students reaching each benchmark | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Advanced | | High | | Intermediate | | Low | |
| 2006 | 8 | (0.5) | 32 | (1.0) | 65 | (1.2) | 87 | (1.0) |
| 2002 | 9 | (0.7) | 39 | (1.3) | 74 | (1.3) | 92 | (0.7) |
| 1998 | 12 | (1.4) | 38 | (2.3) | 68 | (2.4) | 87 | (1.6) |
| 1994 | 11 | (1.2) | 35 | (1.8) | 66 | (1.8) | 85 | (1.7) |

*Note: Standard errors are presented in parentheses.*

## Trends on the test questions

At the end of each cycle of TIMSS, test questions are released into the public domain. At the beginning of the next cycle, new questions are developed to replace the released questions. In addition, in order to provide a trend measure over time, each cycle of TIMSS includes some questions from the previous cycle(s). This section presents an analysis of the trend questions included in both TIMSS 2002/03 and TIMSS 2006/07. Note that no questions from TIMSS 1994/95 were included in the TIMSS 2006/07 assessment.

There were 75 questions common to both the 2002/03 and 2006/07 cycles. Of these 75 questions, 9 questions had similar proportions of students correctly answering them across the two cycles (as shown in Table 3). There were quite a number of questions (45) that proportionally fewer students correctly answered in 2006 compared with 2002. In contrast, there were 21 questions that proportionally more students correctly answered in 2006 compared with 2002. When the change in proportions of students correctly answering was averaged across all the common questions, this represented a decrease of 2 percent.

While this analysis demonstrates a fairly small decrease overall, compared to the decrease of 18 scale score points, it should be remembered that the scale scores are calculated across all countries. Although New Zealand Year 5 students have performed about the same when averaged across questions common to the two cycles, relative to other countries they have decreased significantly between 2002 and 2006.

**Table 3** **Trends in the proportions of students correctly answering science questions common to 2002/03 and 2006/07**

| Change between 2002/03 and 2006/07 | Decrease by 5% or more | Decrease by between 1% and 5% | Increase or decrease by 1% or less | Increase by between 1% and 5% | Increase by 5% or more |
|---|---|---|---|---|---|
| Number of questions | 21 | 24 | 9 | 17 | 4 |

It is interesting to note that of the 21 questions in the group that decreased by 5 percent or more (when the proportion of students correctly answering in 2006 was compared with 2002), there were proportionally more of the physical and earth science questions than life science questions. In contrast, proportionally more life science questions were in the group where proportionally more students answered correctly and far fewer physical science questions.

## Trends in science content and cognitive domains

The science assessment in TIMSS is organised around two dimensions, a content dimension and a cognitive dimension, as described in the *TIMSS 2007 assessment frameworks* (Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberber, 2005). The content dimension comprises three content domains that describe the subject matter to be assessed:

- life science;
- physical science; and
- earth science.

The life science domain is similar to the Living World strand in the New Zealand curriculum and the earth science domain is similar to the Planet Earth and Beyond strand. The physical science domain encompasses both the Material and Physical World strands of the New Zealand curriculum.

The cognitive dimension comprises three cognitive domains that describe the thinking processes that students must use as they engage with the content:

- knowing;
- applying; and
- reasoning.

TIMSS assessment questions were categorised by the content and cognitive domains, and content and cognitive achievement scales were constructed separately for each domain. In order to simplify comparisons across domains, the scales were constructed to have the same average difficulty (set at 500 scale score points). As well as looking at achievement in each of these domains, the results can be used to ascertain relative strengths for participating countries.

As Table 4 shows, New Zealand Year 5 students achieved relatively better at earth science questions and relatively worse at physical science questions in 2006. This is the same pattern as observed in TIMSS 2002/03 (see Caygill, Sturrock, & Chamberlain, 2007). However the differences are more exaggerated in 2006, with a difference between earth science (the highest) and physical science (the lowest) of 17 scale score points. In comparison, in 2002 the difference between earth science (522) and physical science (516) was 6 scale score points.

In the cognitive domains, New Zealand Year 5 students achieved relatively better at tasks that required them to demonstrate their knowledge and relatively worse at questions that required them to apply their knowledge. Year 5 mean science scores on the cognitive domains were not investigated in 2002 so it is not possible to present trend comparisons.

**Table 4 Year 5 mean science scores on the content and cognitive domains in 2006**

| Content domain | Mean domain score | Cognitive domain | Mean domain score |
|---|---|---|---|
| Life science | 506 (2.5) | Knowing | 511 (2.5) |
| Physical science | 498 (2.5) | Applying | 500 (2.1) |
| Earth science | 515 (2.6) | Reasoning | 505 (2.9) |

*Note: Standard errors are presented in parentheses.*

Table 5 shows the number of test questions (and the associated raw score points) in each of the content and cognitive domains. As can be seen from the table, score points were not evenly distributed across domains. This distribution of questions across domains reflects the content and cognitive emphasis of many of the curricula of participating countries.

Looking at Tables 4 and 5 together, it is important to note that the content domain where New Zealand Year 5 students show the greatest strength, earth science, had the least number of questions. In contrast, the cognitive area of greatest strength, knowing, had the greatest number of questions. The distribution of science questions across the content domains was very similar in 2006 to 2002.

**Table 5 Number of questions in each of the content and cognitive domains**

| Content domain | Total number of questions | Total number of score points | Cognitive domain | Total number of questions | Total number of score points |
|---|---|---|---|---|---|
| Life science | 74 | 85 | Knowing | 77 | 89 |
| Physical science | 64 | 67 | Applying | 63 | 68 |
| Earth science | 36 | 42 | Reasoning | 34 | 37 |
| Total | 174 | 194 | Total | 174 | 194 |

*Note: In scoring the tests, correct answers to most questions were awarded one point. However, responses to some constructed-response questions were evaluated for partial credit with a fully correct answer awarded two points. Thus, the number of score points exceeds the number of questions in the test.*

## New Zealand science achievement in 2006 in an international context

As shown in Figure 2, the mean science score for New Zealand Year 5 students in TIMSS 2006/07 was 504 scale score points. New Zealand's mean score was similar to Scotland (500) and significantly higher than 13 other countries. In contrast, 21 countries had higher mean science achievement, including Singapore (587), England (542), the United States (539) and Australia (527).

The range of achievement (from the 5th to 95th percentile) in New Zealand was 299 score points from 344 (the 5th percentile) to 643 (the 95th percentile). This was relatively wider than many of the higher-achieving countries, but narrower than that of Singapore (309). Another measure of spread, the inter-quartile range (from the 25th to 75th percentile) can also be examined. For New Zealand (121) this was wider than many of the higher-performing countries, but the same as Singapore (also 121).

Given the number of countries now participating in TIMSS, it is more meaningful to compare New Zealand to a selection of countries (such as English-speaking or high-performing). Compared to the other countries that tested in English (Singapore, England, the United States, Australia, and Scotland), New Zealand had significantly lower science achievement, on average, than all of them except for Scotland.

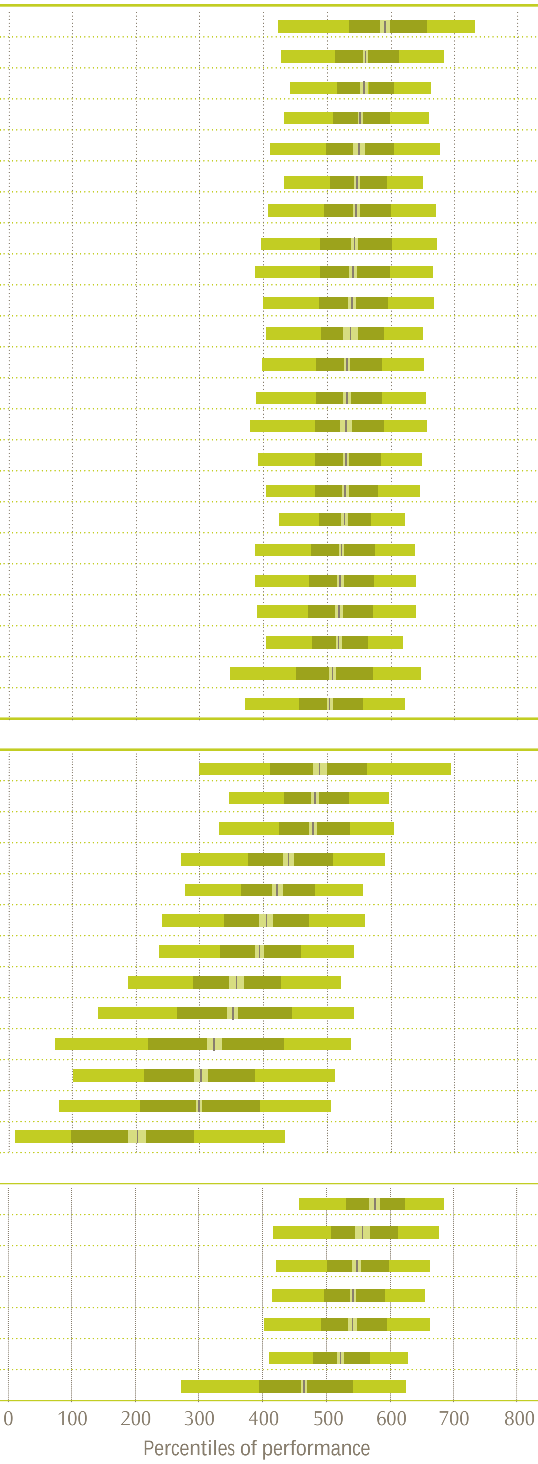
Alongside Figure 2, Table 6 presents some information to help put science achievement in context. Countries are presented in the same order as in Figure 2. It contains information on the number of years of primary schooling students will have undertaken by the time of the assessment, along with students' average age at the time of testing. Also given in the table is the average number of hours of time spent in science instruction during the assessment year according to teacher reports. Three bits of information are presented about the economic circumstances, on average, across each country, the Human Development Index, and two versions of the Gross National Income per Capita (described later).

New Zealand spends less time at the middle primary level teaching science, according to teacher reports, than most of the English-speaking countries and many of the high-performing countries. Note that teachers were asked first whether they taught science as a separate subject and if not, to estimate the amount of time it was taught as part of an integrated programme. Teachers who reported a similar amount of time on average as New Zealand teachers included Australian and Latvian, while teachers in the Russian Federation reported relatively less time in science instruction. Of particular note when examining the context in which to interpret science achievement results, New Zealand teachers are reporting far fewer hours spent in science instruction for the year (21 fewer hours on average) compared with 2002.

Table 6 also presents the Human Development Index (HDI) provided by the United Nations Development Programme (UNDP – for details see *Human Development Report 2007/2008*, p. 229-232). This index was included by Martin, Mullis and Foy (2008) in the international reporting to provide some context around the economic and educational development of TIMSS participating countries. The index ranges from a minimum value of 0 to a maximum value of 1, with high values indicating that people in a country generally enjoy long life expectancy, high levels of school enrolment and adult literacy, and a good standard of living as measured by per capita GDP. New Zealand was relatively high on this scale with a value of 0.943, similar to that of Italy (0.941), and England and Scotland (0.946 – this value is actually for the United Kingdom as no disaggregated data is available for England and Scotland) and lower than that of Australia (0.962) and the United States (0.951).

Perhaps easier to relate to than the HDI, two versions of the Gross National Income (GNI) per Capita are also presented in Table 6. The first of the two columns gives the GNI per Capita in United States dollars while the second is an adjusted value that takes account of comparative purchasing power between each country and the United States. Compared to the countries that assessed in English, New Zealand has the lowest income regardless of which of these values is used.

**Figure 2** **Distribution of middle primary science achievement in TIMSS 2006/07**




| | Country | Average scale score | | Range 5th to 95th percentile | Inter-quartile range 25th to 75th percentile |
|---|---|---|---|---|---|
| | Singapore | 587 (4.1) | ▲ | 309 | 121 |
| | Chinese Taipei | 557 (2.0) | ▲ | 256 | 101 |
| | Hong Kong SAR | 554 (3.5) | ▲ | 222 | 90 |
| | Japan | 548 (2.1) | ▲ | 227 | 90 |
| | Russian Federation | 546 (4.8) | ▲ | 265 | 106 |
| 1 | Latvia | 542 (2.3) | ▲ | 263 | 105 |
| | England | 542 (2.9) | ▲ | 217 | 90 |
| 2* | United States | 539 (2.7) | ▲ | 276 | 113 |
| | Hungary | 536 (3.3) | ▲ | 278 | 110 |
| | Italy | 535 (3.2) | ▲ | 269 | 106 |
| 1 | Kazakhstan | 533 (5.6) | ▲ | 246 | 99 |
| | Germany | 528 (2.4) | ▲ | 254 | 103 |
| | Australia | 527 (3.3) | ▲ | 267 | 105 |
| | Slovak Republic | 526 (4.8) | ▲ | 256 | 103 |
| | Austria | 526 (2.5) | ▲ | 276 | 108 |
| | Sweden | 525 (2.9) | ▲ | 242 | 97 |
| ** | Netherlands | 523 (2.6) | ▲ | 196 | 81 |
| | Slovenia | 518 (1.9) | ▲ | 251 | 100 |
| * | Denmark | 517 (2.9) | ▲ | 253 | 102 |
| | Czech Republic | 515 (3.1) | ▲ | 249 | 100 |
| 1 | Lithuania | 514 (2.4) | ▲ | 214 | 86 |
| | New Zealand | 504 (2.6) | | 299 | 121 |
| * | Scotland | 500 (2.3) | | 252 | 100 |
| | TIMSS scale avg. | 500 | | | |
| | Armenia | 484 (5.7) | ▼ | 396 | 152 |
| | Norway | 477 (3.5) | ▼ | 250 | 101 |
| | Ukraine | 474 (3.1) | ▼ | 274 | 111 |
| | Iran, Islamic Rep. of | 436 (4.3) | ▼ | 320 | 135 |
| 1 | Georgia | 418 (4.6) | ▼ | 279 | 116 |
| | Colombia | 400 (5.4) | ▼ | 319 | 132 |
| | El Salvador | 390 (3.4) | ▼ | 306 | 127 |
| | Algeria | 354 (6.0) | ▼ | 334 | 138 |
| ▶ | Kuwait | 348 (4.4) | ▼ | 401 | 179 |
| | Tunisia | 318 (5.9) | ▼ | 465 | 214 |
| | Morocco | 297 (5.9) | ▼ | 410 | 174 |
| | Qatar | 294 (2.6) | ▼ | 427 | 190 |
| | Yemen | 197 (7.2) | ▼ | 425 | 193 |
| Benchmarking participants | | | | | |
| 2 | Massachusetts, US | 571 (4.3) | ▲ | 228 | 92 |
| 2* | Minnesota, US | 551 (6.1) | ▲ | 260 | 104 |
| 2 | Alberta, C | 543 (3.8) | ▲ | 242 | 98 |
| 2 | British Columbia, C | 537 (2.7) | ▲ | 241 | 95 |
| 2 | Ontario, C | 536 (3.7) | ▲ | 261 | 103 |
| 2 | Quebec, C | 517 (2.7) | ▲ | 218 | 89 |
| ▶ | Dubai, UAE | 460 (2.8) | ▼ | 353 | 147 |


▲ *Country average significantly higher than New Zealand average*

▼ *Country average significantly lower than New Zealand average*

*Note:* * *Met guidelines for sample participation rates only after replacement schools were included.*
*** Nearly satisfied guidelines for sample participation rates only after replacement schools were included.*
*1 National Target Population does not include all of the International Target Population defined by TIMSS.*
*2 National Defined Population covers 90% to 95% of National Target Population.*
▶ *Kuwait and Dubai, UAE tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.*
*Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.*

*Source: Adapted from Exhibits 1.1 and D.1 Martin, Mullis and Foy, 2008.*

**Table 6** **Selected contextual factors for TIMSS 2006/07 countries**

| Country | Years of formal schooling* | Average age at time of testing | Human Development Index** | Gross National Income per capita (in US dollars)*** | GNI per capita (purchasing power parity)*** | Average hours of instructional time in science (teacher reports) |
|---|---|---|---|---|---|---|
| Singapore | 4 | 10.4 | 0.922 | 28730 | 43300 | 82 (0.9) |
| Chinese Taipei | 4 | 10.2 | 0.932 | 17294 | - | 79 (1.6) |
| Hong Kong SAR | 4 | 10.2 | 0.937 | 29040 | 39200 | 72 (5.2) |
| Japan | 4 | 10.5 | 0.953 | 38630 | 32840 | 82 (1.2) |
| Russian Federation | 4 | 10.8 | 0.813 | 5770 | 12740 | 40 (1.1) |
| Latvia | 4 | 11.0 | 0.855 | 8100 | 14840 | 48 (1.2) |
| England | 5 | 10.2 | 0.946 | 40560 | 33650 | 70 (1.7) |
| United States | 4 | 10.3 | 0.951 | 44710 | 44070 | 89 (2.5) |
| Hungary | 4 | 10.7 | 0.874 | 10870 | 16970 | 54 (1.5) |
| Italy | 4 | 9.8 | 0.941 | 31990 | 28970 | 68 (1.4) |
| Kazakhstan | 4 | 10.6 | 0.794 | 3870 | 8700 | 52 (1.3) |
| Germany | 4 | 10.4 | 0.935 | 36810 | 32680 | 106 (2.1) |
| Australia | 4 | 9.9 | 0.962 | 35860 | 33940 | 46 (2.2) |
| Slovak Republic | 4 | 10.4 | 0.863 | 9610 | 17060 | 59 (0.7) |
| Austria | 4 | 10.3 | 0.948 | 39750 | 36040 | 92 (1.1) |
| Sweden | 4 | 10.8 | 0.956 | 43530 | 34310 | 56 (2.5) |
| Netherlands | 4 | 10.2 | 0.953 | 43050 | 37940 | 33 (1.5) |
| Slovenia | 4 | 9.8 | 0.917 | 18660 | 23970 | 84 (0.8) |
| Denmark | 4 | 11.0 | 0.949 | 52110 | 36190 | 59 (0.9) |
| Czech Republic | 4 | 10.3 | 0.891 | 12790 | 20920 | 41 (1.3) |
| Lithuania | 4 | 10.8 | 0.862 | 7930 | 14550 | 51 (0.6) |
| New Zealand | 4.5 - 5.5 | 10.0 | 0.943 | 26750 | 25750 | 45 (2.5) |
| Scotland | 5 | 9.8 | 0.946 | 40560 | 33650 | 51 (3.1) |
| Armenia | 4 | 10.6 | 0.775 | 1920 | 4950 | 81 (4.0) |
| Norway | 4 | 9.8 | 0.968 | 68440 | 50070 | 44 (1.9) |
| Ukraine | 4 | 10.3 | 0.788 | 1940 | 6110 | 33 (1.1) |
| Iran, Islamic Rep. of | 4 | 10.2 | 0.759 | 2930 | 9800 | 83 (2.4) |
| Georgia | 4 | 10.1 | 0.754 | 1580 | 3880 | 35 (2.8) |
| Colombia | 4 | 10.4 | 0.791 | 3120 | 6130 | 139 (3.9) |
| El Salvador | 4 | 11.0 | 0.735 | 2680 | 5610 | 135 (3.5) |
| Algeria | 4 | 10.2 | 0.733 | 3030 | 5940 | 67 (4.7) |
| Kuwait | 4 | 10.2 | 0.891 | 30630 | 48310 | x x |
| Tunisia | 4 | 10.2 | 0.766 | 2970 | 6490 | 71 (2.7) |
| Morocco | 4 | 10.6 | 0.646 | 2160 | 3860 | 54 (4.2) |
| Qatar | 4 | 9.7 | 0.875 | - | - | x x |
| Yemen | 4 | 11.2 | 0.508 | 760 | 2090 | 83 (5.7) |

*Note:* ** Represents years of schooling counting from the first year of primary schooling.*
*** Taken from United Nations Development Programme's Human Development Report 2007/2008. See Martin, Mullis, and Foy for details.*
**** Data on GNI taken from the World Bank's 2008 World Development Indicators. Purchasing Power Parity adjusts the GNI to take account of comparative purchasing power between the country and the United States.*
*Standard errors are presented in parentheses.*

*Source: Adapted from Exhibits 3, 1.1, and 5.3, Martin, Mullis, and Foy, 2008.*

## International trends in science achievement at the middle primary level

There are several ways that trends since 1994 can be examined for the countries participating in TIMSS. The analyses presented here will include only those countries that have participated in all three international cycles, 1994/95, 2002/03, and 2006/07. Table 7 shows the change in mean science scores since 1994/95, ordered so that those countries that have had the biggest positive change since the first cycle are at the top and those with the biggest negative change are at the bottom.

**Table 7 Trends in middle primary school mean science achievement over three cycles of TIMSS**

| Country | 1994/95 to 2006/07 difference | | 2002/03 to 2006/07 difference | |
|---|---|---|---|---|
| Singapore | 63 (6.4) | ▲ | 22 (6.8) | ▲ |
| Latvia | 56 (5.4) | ▲ | 12 (3.5) | ▲ |
| Iran, Islamic Rep. of | 55 (6.3) | ▲ | 22 (5.9) | ▲ |
| Slovenia | 54 (3.6) | ▲ | 28 (3.2) | ▲ |
| Hong Kong SAR | 46 (4.8) | ▲ | 12 (4.6) | ▲ |
| Hungary | 28 (4.8) | ▲ | 6 (4.5) | |
| England | 14 (4.2) | ▲ | 1 (4.4) | |
| Australia | 6 (4.9) | | 7 (5.3) | |
| New Zealand | -1 (5.9) | | -18 (3.5) | ▼ |
| United States | -3 (4.3) | | 3 (3.5) | |
| Japan | -5 (2.6) | ▼ | 4 (2.5) | |
| Netherlands | -7 (4.0) | | -2 (3.1) | |
| Scotland | -14 (5.0) | ▼ | -2 (3.6) | |
| Norway | -27 (5.2) | ▼ | 10 (3.5) | ▲ |

*Note:* ▲ *2006/07 score significantly higher.*
▼ *2006/07 score significantly lower.*
*Standard errors are presented in parentheses.*
*Source: Adapted from Exhibit 1.3 Martin, Mullis, and Foy, 2008.*

Singapore, Latvia, the Islamic Republic of Iran, Slovenia, Hong Kong SAR, Hungary, and England have all had significant increases in mean science achievement since 1994. Of these countries, Latvia and Slovenia have both made significant changes in their education systems since 1994; summaries of these are presented in the following paragraphs. However first we look at Singapore.

### Singapore

Singapore is the country with the largest change over time in mean science score. Singapore has had some changes in the education system over this time with a new science curriculum introduced in 2001, along with a change in the philosophy of the education system since 2004. Singapore has moved toward

*"more quality in terms of classroom interaction, opportunities for expression, and the learning of life long skills through innovative and effective teaching approaches, and away from quantity in terms of rote learning, repetitive tests, and following prescribed answers and set formulae. It also reaffirms the learner at the center [sic] of all that is being done and better recognizes and caters to the various needs and interests of different learners."* (Quek et al., 2008, p. 537).

Singapore places a great emphasis on the teaching and learning of science and mathematics. Primary schools have science rooms and sometimes mathematics rooms that serve as a focal point for science and mathematics (respectively) activities and innovations. From upper primary levels onward, students have specialist teachers in science and mathematics. Guidelines are given nationally on the number of hours per week mathematics and science should be taught at each grade level.

### Latvia

Since 1998, Latvia has had a basic education standard for students in grades 1 to 9 (aged 7 to 16).[7] Subject standards, which are part of the basic education standard, determine the main aims and tasks of the subject, the mandatory content of the subject, and the forms and order of the evaluation of achievement. The number of lessons per week is set nationally and mandatory. In grades 1 to 4 students have one teacher for all subjects; from grade 5, students have specialist subject teachers. Latvian students have tests in all grades, but the first national assessments occur at the end of grade 3 (students aged 9 and 10 – see Geske, Grinfelds, & Ozola, 2008).

### Slovenia

Slovenia has been undergoing some significant changes in its schooling system, the most obvious of which is the lowering of the school starting age from 7 to 6 and revised national curricular documents for all levels of pre-university education. The goal of the reforms, implementation of which began in 1999, are:

*"a higher level of interconnectedness of disciplinary knowledge, and increased active role of students, internationally comparable standards and levels of knowledge, improvement in functional literacy, and an increase in the quality and longevity of acquired knowledge."* (Japelj Pavešić & Svetlik, 2008, p. 537).

The Slovenian syllabus specifies the exact number of yearly and weekly lessons for individual subjects. In grades 1 to 3, nearly all subjects are taught by general class teachers. During grades 4 to 6, specialist teachers become more and more involved in the teaching process.

### Relative rankings

In many summaries of the international data, relative rankings of mean scores are used to describe change. This is not a particularly desirable practice as any mean scores derived from a sample and ascribed to a population have some level of uncertainty around them and rankings ignore this uncertainty. In addition, some presentations of rankings fail to mention the number of countries included in the ranking.

Table 8 presents relative ranking changes between 1994/95 and 2006/07. This should be read with caution, because, although a country may be ranked higher, the mean scores may not be significantly different when the uncertainties are taken into account. For example, the mean science achievement for Hong Kong SAR and that of Japan in 2006/07 are not significantly different.

Table 8 shows that for New Zealand, not only are the mean science scores essentially the same for 1994 and 2006, but that there is very little movement in our position in the 'league tables' when only those countries in both assessments are included. Despite New Zealand's Year 5 mean science score being the same in 1994/95 as it was in 2006/07, it has moved from being around the mean of the 14 countries to being below the mean over this time period.

7 Pre-primary education is compulsory for students aged 5 and 6 in Latvia.

**Table 8** **Middle primary mean science scores for countries participating in three cycles of TIMSS from 1994/95 to 2006/07**

| 1994/95 mean science score | | | | 2002/03 mean science score | | | | 2006/07 mean science score | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Japan | 553 | (1.8) | ▲ | Singapore | 565 | (5.5) | ▲ | Singapore | 587 | (4.1) | ▲ |
| United States | 542 | (3.3) | ▲ | Japan | 543 | (1.5) | ▲ | Hong Kong SAR | 554 | (3.5) | ▲ |
| Netherlands | 530 | (3.2) | ▲ | Hong Kong SAR | 542 | (3.1) | ▲ | Japan | 548 | (2.1) | ▲ |
| England | 528 | (3.1) | ▲ | England | 540 | (3.6) | ▲ | Latvia | 542 | (2.3) | ▲ |
| Singapore | 523 | (4.8) | ▲ | United States | 536 | (2.5) | ▲ | England | 542 | (2.9) | ▲ |
| Australia | 521 | (3.8) | ▲ | Latvia | 530 | (2.8) | ▲ | United States | 539 | (2.7) | ▲ |
| Scotland | 514 | (4.5) | ▲ | Hungary | 530 | (3.0) | ▲ | Hungary | 536 | (3.3) | ▲ |
| Hong Kong SAR | 508 | (3.3) | | Netherlands | 525 | (2.0) | ▲ | Australia | 527 | (3.3) | |
| Hungary | 508 | (3.4) | | New Zealand | 523 | (2.3) | ▲ | Netherlands | 523 | (2.6) | |
| New Zealand | 505 | (5.3) | | Australia | 521 | (4.2) | | Slovenia | 518 | (1.9) | ▼ |
| Norway | 504 | (3.7) | | Scotland | 502 | (2.9) | ▼ | New Zealand | 504 | (2.6) | ▼ |
| Latvia | 486 | (4.9) | ▼ | Slovenia | 490 | (2.5) | ▼ | Scotland | 500 | (2.3) | ▼ |
| Slovenia | 464 | (3.1) | ▼ | Norway | 466 | (2.6) | ▼ | Norway | 477 | (3.5) | ▼ |
| Iran, Islamic Rep. of | 380 | (4.6) | ▼ | Iran, Islamic Rep. of | 414 | (4.1) | ▼ | Iran, Islamic Rep. of | 436 | (4.3) | ▼ |
| Mean for all 14* | 505 | (1.1) | | Mean for all 14* | 516 | (1.0) | | Mean for all 14* | 524 | (1.1) | |

*Note:* ▲ *Country mean is significantly higher than the mean for the 14 countries*
▼ *Country mean is significantly lower than the mean for the 14 countries*
** This mean has been calculated for the 14 countries common to all cycles. It is calculated by pooling all student results for the 14 countries and weighting so that each country contributes equally to the mean.*
*Standard errors are presented in parentheses.*

## International trends in science benchmarks

As shown in Table 9, eight percent of New Zealand Year 5 students reached the advanced benchmark, the point where students were deemed capable of applying *knowledge and understanding of scientific processes and relationships in beginning scientific inquiry*. This was a similar proportion to countries including Austria (9%), Sweden (8%), and the Czech Republic (7%), and higher than Slovenia (6%), Scotland (4%), and the Netherlands (4%). However, Singapore was the country with the greatest proportion of students at the advanced benchmark, more than four times the proportion of New Zealand students, at 36 percent.

Examining the low benchmark, 13 percent of New Zealand students did not reach this benchmark and therefore, in terms of the benchmark definition, did not demonstrate *some elementary knowledge of life science and physical science*. Most countries had some students in this group, with Hong Kong SAR (2%) and Latvia (2%) having the fewest students unable to reach the low benchmark. Countries with similar proportions at the advanced benchmark generally had fewer students unable to reach the low benchmark when compared to New Zealand.

Included in the table is the international median percentage of students at each benchmark. Approximately the same proportion of New Zealand Year 5 students reached the advanced benchmark as the international median, so New Zealand was around the middle of the countries for the advanced benchmark. For the high, intermediate, and low benchmarks, proportionally fewer New Zealand Year 5 students reached these benchmarks compared to the international median.

**Table 9 Proportion of middle primary students at each international benchmark**

| | Country | Advanced | High | Intermediate | Low |
|---|---|---|---|---|---|
| | | Percentage of students reaching each benchmark | | | |
| | Singapore | 36 (1.9) | 68 (1.9) | 88 (1.1) | 96 (0.5) |
| | Chinese Taipei | 19 (1.0) | 55 (1.2) | 86 (0.7) | 97 (0.4) |
| | Russian Federation | 16 (1.9) | 49 (2.3) | 82 (1.7) | 96 (0.9) |
| 2* | United States | 15 (0.9) | 47 (1.4) | 78 (1.1) | 94 (0.6) |
| | England | 14 (1.2) | 48 (1.6) | 81 (1.1) | 95 (0.6) |
| | Hong Kong SAR | 14 (1.4) | 55 (2.2) | 88 (1.2) | 98 (0.4) |
| | Hungary | 13 (1.0) | 47 (1.8) | 78 (1.6) | 93 (0.8) |
| | Italy | 13 (1.0) | 44 (1.6) | 78 (1.3) | 94 (0.7) |
| | Japan | 12 (1.0) | 51 (1.1) | 86 (1.0) | 97 (0.4) |
| | Armenia | 12 (1.8) | 27 (1.8) | 52 (1.8) | 77 (1.6) |
| | Slovak Republic | 11 (0.8) | 42 (2.0) | 75 (1.8) | 92 (1.3) |
| | Australia | 10 (0.7) | 41 (2.2) | 76 (1.6) | 93 (0.8) |
| 1 | Latvia | 10 (1.1) | 47 (1.7) | 84 (1.3) | 98 (0.4) |
| | Germany | 10 (0.7) | 41 (1.3) | 76 (1.2) | 94 (0.6) |
| 1 | Kazakhstan | 10 (1.3) | 44 (3.1) | 79 (2.6) | 95 (1.0) |
| | Austria | 9 (0.7) | 39 (1.3) | 76 (1.3) | 93 (0.6) |
| | Sweden | 8 (0.6) | 37 (1.6) | 76 (1.5) | 95 (0.6) |
| | New Zealand | 8 (0.5) | 32 (1.0) | 65 (1.2) | 87 (1.0) |
| | Czech Republic | 7 (0.7) | 33 (1.9) | 72 (1.4) | 93 (0.8) |
| * | Denmark | 7 (0.8) | 35 (1.5) | 72 (1.5) | 93 (0.8) |
| | Slovenia | 6 (0.6) | 36 (1.3) | 74 (1.0) | 93 (0.6) |
| * | Scotland | 4 (0.6) | 26 (1.2) | 65 (1.3) | 90 (0.8) |
| ** | Netherlands | 4 (0.8) | 34 (1.8) | 79 (1.4) | 97 (0.5) |
| 1 | Lithuania | 3 (0.4) | 30 (1.4) | 74 (1.4) | 95 (0.6) |
| | Ukraine | 2 (0.3) | 17 (1.1) | 52 (1.5) | 82 (1.3) |
| | Iran, Islamic Rep. of | 2 (0.3) | 12 (1.0) | 36 (1.7) | 65 (1.9) |
| | Norway | 1 (0.4) | 17 (1.4) | 54 (2.0) | 84 (1.4) |
| | Colombia | 1 (0.2) | 6 (0.8) | 22 (1.7) | 51 (2.4) |
| 1 | Georgia | 1 (0.2) | 5 (0.8) | 26 (2.0) | 59 (2.1) |
| | El Salvador | 0 (0.1) | 4 (0.5) | 18 (1.2) | 47 (1.5) |
| ▶ | Kuwait | 0 (0.2) | 4 (0.6) | 16 (1.3) | 37 (1.3) |
| | Morocco | 0 (0.2) | 2 (0.5) | 9 (1.4) | 21 (1.9) |
| | Algeria | 0 (0.1) | 2 (0.5) | 11 (1.3) | 33 (2.1) |
| | Tunisia | 0 (0.1) | 3 (0.6) | 14 (1.1) | 32 (1.7) |
| | Qatar | 0 (0.0) | 2 (0.2) | 8 (0.3) | 23 (0.7) |
| | Yemen | 0 (0.0) | 0 (0.1) | 2 (0.5) | 8 (1.1) |
| International Median | | 7 | 34 | 74 | 93 |
| Benchmarking participants | | | | | |
| 2 | Massachusetts, US | 22 (2.2) | 64 (2.4) | 92 (1.1) | 99 (0.4) |
| | Minnesota, US | 17 (1.9) | 54 (3.2) | 84 (2.1) | 96 (1.5) |
| 2 | Alberta, C | 12 (1.3) | 48 (2.0) | 82 (1.5) | 96 (0.7) |
| 2 | Ontario, C | 12 (1.2) | 45 (2.2) | 79 (1.7) | 95 (1.0) |
| 2 | British Columbia, C | 11 (0.8) | 44 (1.7) | 81 (1.5) | 96 (0.6) |
| 2 | Quebec, C | 5 (0.6) | 32 (1.9) | 74 (1.9) | 96 (0.6) |
| ▶ ** | Dubai, UAE | 4 (0.5) | 21 (1.0) | 48 (1.3) | 72 (1.4) |

*Note:* ** Met guidelines for sample participation rates only after replacement schools were included.*
*** Nearly satisfied guidelines for sample participation rates only after replacement schools were included.*
*1 National Target Population does not include all of the International Target Population defined by TIMSS.*
*2 National Defined Population covers 90% to 95% of National Target Population.*
*▶ Kuwait and Dubai, UAE tested the same cohort of students as other countries, but later in 2007, at the beginning of the next school year.*
*Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.*

*Source:* *Adapted from Exhibit 2.2 Martin, Mullis and Foy, 2008.*

Figures 3 to 6 present examples of questions that Year 5 students achieving at or above the *advanced, high, intermediate*, and *low* benchmarks were likely to have answered correctly. An example of a correct answer and a summary of the scoring guide are presented. In addition, proportions of students successfully completing the question for a selection of countries, including the best performing country on that question, are shown. The international average is also presented as an indication of how students in all 37 countries performed on this question.

### Figure 3 Question students reaching the advanced benchmark are likely to have answered correctly

Question with example of correct answer
Content domain: life science
Cognitive domain: reasoning

There is a giant turtle that lives on an island. He is the only turtle left of a special type of giant turtle.

Can he reproduce so that this type of turtle does not die out?

(Tick one box.)

☐ Yes
☑ No

Give a reason for this answer.

Turtles cannot reproduce all by themselves. It is a male turtle so he needs a female.

Scoring guide

Explains that the last surviving member of a species of a turtle cannot reproduce and gives a reason.

| Country | Percent full credit | |
|---|---|---|
| Lithuania | 58 | (2.4) |
| Australia | 48 | (2.5) |
| England | 47 | (2.4) |
| Japan | 45 | (2.1) |
| Chinese Taipei | 43 | (2.4) |
| United States | 42 | (1.6) |
| Singapore | 38 | (2.4) |
| Hong Kong SAR | 36 | (2.2) |
| Scotland | 36 | (2.1) |
| New Zealand | 35 | (2.0) |
| International Avg. | 30 | (0.3) |

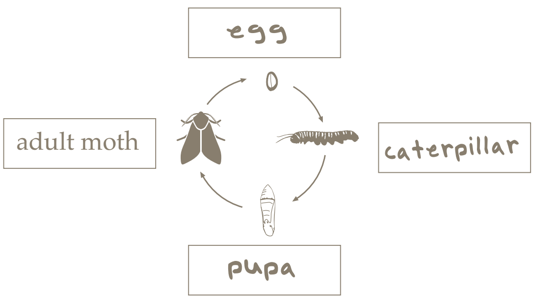
*Note: Standard errors are presented in parentheses.*
*Source: Adapted from Exhibit 2.6, Martin, Mullis, and Foy, 2008.*

### Figure 4 Question students reaching the high benchmark are likely to have answered correctly

Question with example of correct answer
Content domain: life science
Cognitive domain: knowing

The diagram below shows the life cycle of a moth.

Write the name of each stage in the boxes provided. One stage has been completed for you.




Scoring guide

All three of egg; caterpillar or larva; chrysalis, pupa or cocoon.

| Country | Percent full credit | |
|---|---|---|
| Japan | 93 | (1.3) |
| Singapore | 64 | (2.0) |
| Chinese Taipei | 61 | (2.4) |
| Australia | 56 | (2.5) |
| New Zealand | 52 | (1.9) |
| United States | 48 | (1.8) |
| Lithuania | 43 | (2.8) |
| England | 36 | (2.2) |
| Scotland | 33 | (2.5) |
| International Avg. | 33 | (0.4) |
| Hong Kong SAR | 22 | (2.1) |



*Note: Standard errors are presented in parentheses.*
*Source: Adapted from Exhibit 2.9, Martin, Mullis, and Foy, 2008.*

## Figure 5 Question students reaching the intermediate benchmark are likely to have answered correctly

Question with example of correct answer
Content domain: physical science
Cognitive domain: applying

When you blow into water using a straw, bubbles are formed and rise to the top.

Why do the bubbles rise in the water?

They rise because they are made from air which is lighter than water

Scoring guide

Adequately explains why bubbles rise to the top.

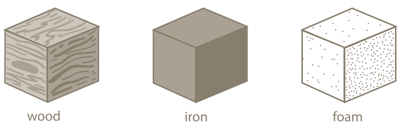
| Country | Percent full credit | |
|---|---|---|
| Russian Federation | 79 | (2.3) |
| Chinese Taipei | 77 | (1.7) |
| Singapore | 72 | (1.9) |
| Australia | 67 | (2.8) |
| England | 66 | (2.3) |
| Japan | 65 | (2.0) |
| New Zealand | 64 | (1.8) |
| United States | 61 | (1.7) |
| Lithuania | 61 | (2.4) |
| Scotland | 54 | (2.4) |
| International Avg. | 51 | (0.4) |
| Hong Kong SAR | 48 | (2.4) |

*Note: Standard errors are presented in parentheses.*
*Source: Adapted from Exhibit 2.12, Martin, Mullis, and Foy, 2008.*

## Figure 6 Question students reaching the low benchmark are likely to have answered correctly

Question with example of correct answer
Content domain: physical science
Cognitive domain: applying

The three objects below are the same shape and size.




Which statement about the weight of the object is most likely to be correct?

- Ⓐ The wood object is the heaviest.
- ● The iron object is the heaviest.
- Ⓒ The foam object is the heaviest.
- Ⓓ All three objects weigh the same.

Scoring guide

Selects multiple-choice answer B

| Country | Percent full credit | |
|---|---|---|
| Japan | 94 | (1.2) |
| Chinese Taipei | 91 | (1.4) |
| Hong Kong SAR | 90 | (1.4) |
| England | 89 | (1.6) |
| Singapore | 88 | (1.4) |
| Scotland | 82 | (1.8) |
| United States | 80 | (1.1) |
| International Avg. | 80 | (0.3) |
| Australia | 68 | (3.1) |
| New Zealand | 67 | (2.3) |

*Note: Standard errors are presented in parentheses.*
*Source: Adapted from Exhibit 2.15, Martin, Mullis, and Foy, 2008.*

## International trends in science content and cognitive domains

As mentioned earlier, New Zealand Year 5 students demonstrated a relative strength in earth science questions and a relative weakness in physical science questions. Relatively higher earth science mean scores and relatively lower physical science mean scores were also observed for Scotland and Australia. In contrast, the higher-achieving countries, Singapore, Chinese Taipei, and Japan all showed a relative strength in the physical science domain.

In the cognitive domains, New Zealand Year 5 students demonstrated a relative strength in questions that required them to demonstrate their knowledge and a relative weakness in questions that required them to apply their knowledge. This pattern was the same as Scotland and the United States. In contrast, Chinese Taipei, Hong Kong SAR, and Japan all showed a relative strength in the reasoning domain. Interestingly, Singaporean students were relatively stronger at questions in the knowing domain and relatively weaker at the questions in the reasoning domain.

## TIMSS and the New Zealand science curriculum

### Science curriculum levels and the TIMSS content domains

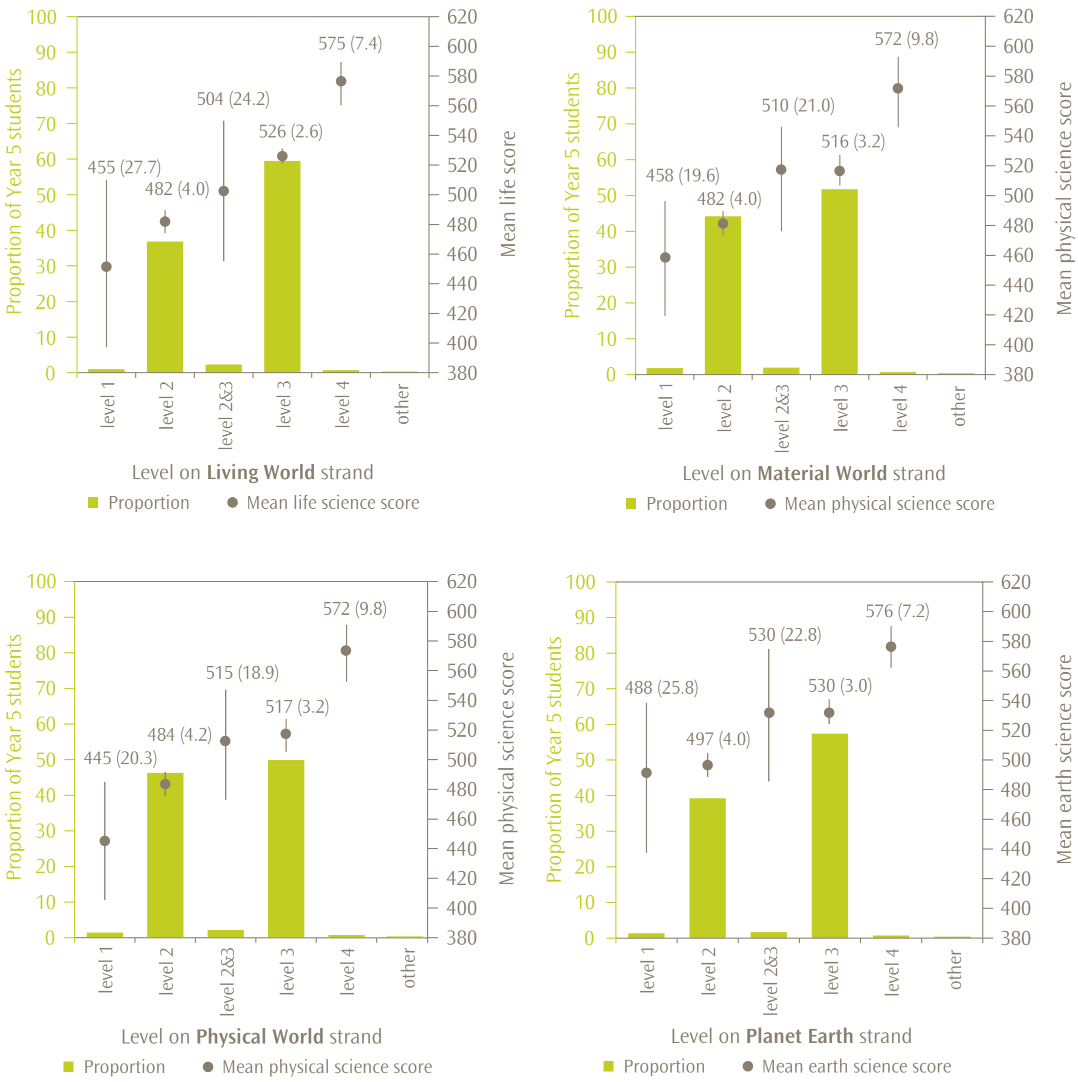
In order to gain greater understanding of the relationship between the *Science in the New Zealand Curriculum* (SciNZC)[8] levels and student achievement in TIMSS, New Zealand teachers were asked at which level of the SciNZC most of the students in their class were currently working for each of the strands: *Living World, Material World, Physical World, Planet Earth and Beyond*. Note that the information was not collected for individual students but for the majority of the TIMSS students the teacher taught. For the purpose of analysis, a teacher's response has been assigned to each individual student in the class. Figure 7 shows that while the majority of Year 5 students were working at level 3 of the curriculum, there were still a significant number of students in classes working at level 2.

Since the TIMSS domains were similar to the SciNZC strands in terms of content, with *Material World* and *Physical World* combined similar to the *physical science* domain, the figure also maps the mean TIMSS domain score for the students estimated to be working at each level of the SciNZC. For example, the mean score in the TIMSS *life science* domain for the 59 percent of students in classes estimated to be working at level 3 of the *Living World* strand of the SciNZC was 526 scale score points. The figure shows that students whose classes were working at higher levels of the curriculum have higher achievement on the associated TIMSS content domain. Note that no attempt is being made here to infer a causal link – that is, we are not saying the higher mean achievement is because they are working at the higher level.

It is interesting to look at these results in an international context and observe that if only those students working at level 3 of the curriculum were included in the TIMSS testing, New Zealand's overall science score would still have been below that of the high-performing countries, Singapore, Chinese Taipei, Hong Kong SAR, and Japan. For example, the mean score for Singaporean students on the life science domain was 582 scale score points, while New Zealand students working at level 3 had a mean score of 526 scale score points.

8 This was the curriculum in place at the time of testing.

**Figure 7 Mean content area achievement by New Zealand curriculum strands**




Note: *The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. In the cases where there are no mean scores, the 'other' grouping, there were too few students to report achievement. Scores presented for levels 1, 4, and combined 2 & 3 should be treated with caution as the proportion of students is in each of these groups is small. Standard errors are presented in parentheses.*

## Curriculum match

Questions about international studies often focus on the appropriateness of the assessment questions for New Zealand students. New Zealand is not unique in asking this question; other countries are also concerned with appropriateness of the tests. The TIMSS assessment questions are developed through a collaborative process that begins with the development of an assessment framework. The *TIMSS 2007 assessment frameworks* (Mullis, Martin, et al., 2005) were designed to specify the important aspects of science that participating countries agreed should be the focus of an international assessment of science achievement. However it is inevitable that the tests included questions that were unfamiliar to some students in some countries. In order to investigate the extent to which the TIMSS 2006/07 assessment was relevant to each country's curriculum, TIMSS conducted a Test-Curriculum Matching Analysis (TCMA). The TCMA was also used to investigate the impact of selecting only appropriate questions on a country's performance.

For the TCMA, each question was examined using the following two criteria:

- whether or not the topic of the question is in the intended curriculum for the majority of middle primary students (in our case Year 5) – that is, more than 50 percent; and
- whether or not the question topic is intended to be encountered by the middle primary students prior to the TIMSS testing (testing of New Zealand Year 5 students occurred in the beginning of November).

While all questions, regardless of this analysis, were included in any overall results reported for TIMSS, this analysis was used to ascertain the level to which the results might change for New Zealand if only questions judged appropriate were included in the tests. The analysis also included an examination of how students in other countries would fare if given only the "New Zealand-appropriate" test.

Table 10 shows the proportion of questions considered appropriate to the New Zealand curriculum in each of the TIMSS content areas. However, it should be noted that New Zealand's science curriculum provides some challenges for deciding whether at least half Year 5 students are likely to have met the question topics in the TIMSS test.[9] The curriculum is not prescriptive, instead providing some broad guidelines of science concepts and skills that schools can choose to cover. Schools are encouraged to design science programmes that are relevant to their students and communities. Consequently, when schools plan their science programmes there is considerable variation between them. Another challenge is that the broad achievement objectives are grouped in Levels which cover approximately two years of schooling. As shown in the previous section, New Zealand Year 5 students were generally working at levels 2 and 3 of the curriculum so information from levels 1, 2, and 3 was used to guide judgements on the TCMA.[10]

**Table 10 Appropriateness of the TIMSS tests to the New Zealand curriculum**

| TIMSS content domain | Number of score points judged appropriate for New Zealand curriculum | Number of score points in TIMSS assessment | Proportion of score points judged appropriate for New Zealand curriculum |
|---|---|---|---|
| Life science | 63 | 85 | 74% |
| Physical science | 44 | 66 | 67% |
| Earth science | 30 | 42 | 71% |

*Note: Life science corresponds to the Living World strand of the curriculum, physical science corresponds to a combination of the Physical and Material World strands of the curriculum, and earth science corresponds to the Planet Earth and Beyond strand of the curriculum.*

Although only around two-thirds of the questions were judged appropriate for New Zealand students, the TIMSS TCMA analysis shows that some of the higher-performing countries would have done better on the 'New Zealand' test than New Zealand Year 5 students, as shown in Table 11.

9 Note that for the TCMA, the curriculum document used was the 1993 Science in the New Zealand Curriculum (Ministry of Education, 1993).

10 Thanks to Chris Joyce and Ally Bull from NZCER for this work.

**Table 11 Performance of middle primary students from selected countries on the 'New Zealand' appropriate test**

| Country | Average percent correct on New Zealand test |
|---|---|
| Singapore | 71 |
| Chinese Taipei | 63 |
| Hong Kong SAR | 63 |
| Russian Federation | 62 |
| Japan | 60 |
| England | 60 |
| United States | 61 |
| Australia | 58 |
| Scotland | 53 |
| New Zealand | 53 |

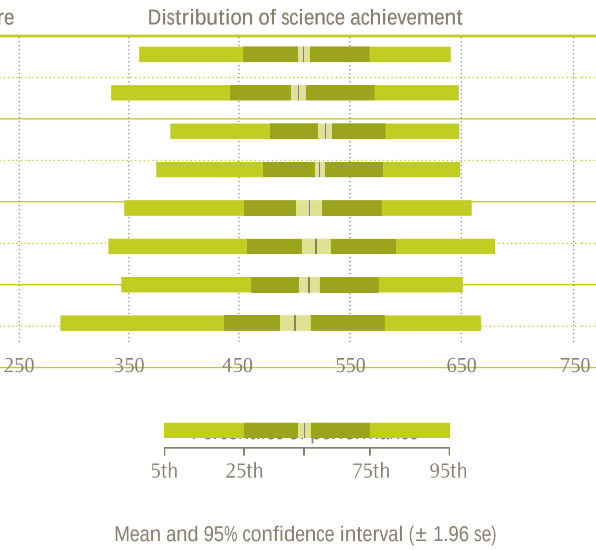
*Source: Adapted from Exhibit C.1 in Martin, Mullis, and Foy, 2008.*

## Science achievement by gender

There was no significant difference in mean science achievement between Year 5 boys (502) and girls (506) in 2006. However, the distribution of achievement was wider for boys (313) than for girls (281) as shown in Figure 8. As Figure 8 also shows, the wider distribution among boys has been consistent over the four cycles. Similarly, there was no significant difference between the mean science achievement of boys and that of girls in each of the preceding cycles, with the exception of 1994 when girls had higher mean performance than boys. It is interesting to observe the narrowing of the distribution of science achievement for boys between 1994 and 2006.

**Figure 8 Trends in distributions of achievement for girls and boys from 1994 to 2006**

| Year | | Mean science score | Range from 5th to 95th percentile | Inter-quartile range from 25th to 75th percentile |
|---|---|---|---|---|
| 2006 | Girls | 506 (2.8) | 281 | 114 |
| | Boys | 502 (3.5) | 313 | 130 |
| 2002 | Girls | 526 (3.2) | 260 | 104 |
| | Boys | 521 (2.3) | 274 | 108 |
| 1998 | Girls | 511 (5.9) | 313 | 124 |
| | Boys | 518 (6.6) | 348 | 134 |
| 1994 | Girls | 511 (4.8) | 308 | 115 |
| | Boys | 499 (7.0) | 379 | 145 |




*Note: Standard errors are presented in parentheses.*

## Benchmarks for Year 5 boys and girls

There was little difference between the proportions of boys and girls reaching each of the international benchmarks except for the low benchmark, where the difference was statistically significant (see Table 12). Fewer boys (85%) reached the low benchmark compared with the girls (89%). Another way of looking at this is to examine the proportions of boys and girls who did not reach this low benchmark. Proportionally more boys (15%) than girls (11%) did not reach the low benchmark – in terms of the benchmark definitions, these students did not demonstrate some elementary knowledge of life science and physical science.[11]

**Table 12: Proportion of Year 5 students reaching each international benchmark by gender in TIMSS 2006/07**

| | Percentage of Year 5 students reaching each international benchmark | | | |
|---|---|---|---|---|
| Gender | Advanced | High | Intermediate | Low |
| girls | 7 (0.7) | 31 (1.4) | 67 (1.6) | 89 (0.9) |
| boys | 9 (0.9) | 33 (1.6) | 63 (1.5) | 85 (1.3) |

*Note: Standard errors are presented in parentheses.*

Proportionally more boys and girls were lower achievers in 2006 compared with 2002, that is, did not reach the low benchmark. In 2002, seven percent of girls and eight percent of boys did not reach this benchmark (see Caygill, Sturrock, & Chamberlain, 2007, p. 68). This pattern was also observed across the low, intermediate, and high benchmarks, with proportionally fewer girls and boys reaching each of these benchmarks in 2006 compared with 2002. However, there was no significant difference between the proportions of girls and boys reaching the advanced benchmark in 2006 compared with 2002. When comparing 1994 and 2006, there were no significant differences in the proportion of girls and boys at each of the benchmarks.

## Achievement on the content and cognitive domains for girls and boys

While there were no overall differences in mean science achievement between girls and boys, there were some distinct differences in terms of the content and cognitive domains. On average, girls had higher scores in life science, while boys had higher scores in earth science (see Table 13). Boys and girls performed similarly in the physical science domain. Boys and girls also performed similarly on questions involving knowledge or applying that knowledge. However, girls on average performed better than boys on questions involving reasoning.

**Table 13 Year 5 mean science scores on the content and cognitive domains by gender**

| | Mean domain score | |
|---|---|---|
| Content domain | girls | boys |
| Life science | 512 (3.0) ▲ | 501 (3.8) |
| Physical science | 500 (3.2) | 497 (3.2) |
| Earth science | 512 (2.9) | 518 (3.0) ▲ |

| | Mean domain score | |
|---|---|---|
| Cognitive domain | girls | boys |
| Knowing | 513 (3.1) | 508 (3.1) |
| Applying | 498 (2.7) | 501 (3.2) |
| Reasoning | 514 (3.1) ▲ | 497 (4.0) |

*Note: ▲ mean domain score significantly higher than other gender.*
*Standard errors are presented in parentheses.*
*Source: Exhibit 3.3 from Martin, Mullis & Foy, 2008.*

11 See the *Trends in benchmarks for science* section earlier in this report for the full definition.

## Science achievement by ethnicity, language, and country of birth

This section will examine the science achievement of students in TIMSS across different ethnic groups, by use of English at home, and by country of birth. These three characteristics of students are interrelated so in the final part of this section they are examined together. This section will examine relationships with science achievement, but it should be noted that the existence of a relationship does not infer a causal link.

### Science achievement by ethnicity

Five broad ethnic classifications are used to describe ethnicity in New Zealand. They are: Päkehä/European, Mäori, Pasifika, Asian, and 'Other' ethnic groupings. The majority of Year 5 students in New Zealand were identified by their schools[12] as Päkehä/European (61%) or Mäori (19%). Pasifika (10%) and Asian (7%) students made up most of the rest of the ethnic groupings, with four percent of students categorised in the Other ethnic grouping.

Previous cycles of TIMSS have shown that average science achievement varies across ethnic groups. Although the variation in achievement is not caused by ethnicity *per se*, education policies have been introduced so that all students may realise their potential. Specific areas of focus for the Ministry of Education include the achievement of Mäori and Pasifika students (Ministry of Education, 2007). The results at the Year 5 level in TIMSS 2002/03 (Caygill, Sturrock & Chamberlain, 2007) showed an increase in science performance, on average, for Mäori and Pasifika students since the first cycle in 1994/1995.

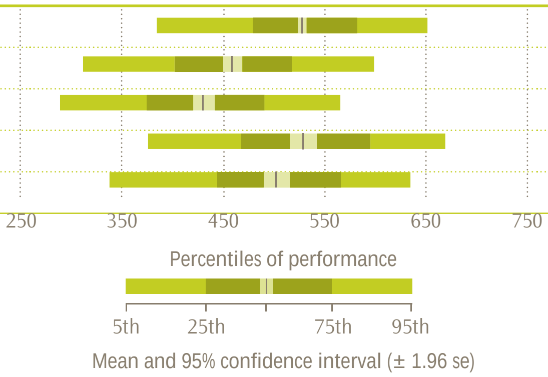
In TIMSS 2006/07, Asian (529) and Päkehä/European (528) students had significantly higher mean science achievement than did their Mäori (459), Pasifika (431) and Other (502) counterparts, as shown in Figure 9. On average, Mäori students performed significantly higher in science than Pasifika students. No significant difference was observed between Päkehä/European and Asian students.

The distribution of achievement of Asian students and those in the Other ethnic grouping was the widest, while the distribution for the Päkehä/European students was the narrowest. Note that the 5th and 95th percentiles of achievement for the students in the Pasifika, Asian, and Other ethnic grouping should be treated with caution as there are few students at these ends of the distribution due to the smaller number of students in these ethnic groupings overall.

12 Based on enrolment information supplied by parents.

**Figure 9** **Distribution of New Zealand Year 5 science achievement for each ethnic grouping**

| Ethnic grouping | Mean science score | Distribution of science achievement | Range from 5th to 95th percentile | Inter-quartile range from 25th to 75th percentile |
|---|---|---|---|---|
| Päkehä/European | 528 (2.3) | | 267 | 103 |
| Mäori | 459 (4.9) | | 287 | 116 |
| Pasifika | 431 (5.4) | | 276 | 116 |
| Asian | 529 (6.8) | | 293 | 127 |
| Other | 502 (6.7) | | 297 | 122 |




*Note: The distribution and ranges for the students in the Pasifika, Asian, and Other ethnic groupings should be read with caution as there is a lot of uncertainty at the extremes of the distribution.*
*Standard errors are presented in parentheses.*

### Benchmarks for ethnic groupings

Within all ethnic groupings, there were students who reached the advanced benchmark; in terms of the benchmark definitions, they demonstrated the ability to complete tasks requiring applying *knowledge and understanding of scientific processes and relationships in beginning scientific inquiry*. Similarly, within all ethnic groupings there were students who did not reach the low benchmark; that is, they did not demonstrate the ability to complete a reasonable number of the simplest science tasks which TIMSS seeks to measure.

Higher proportions of Asian and Päkehä/European students reached the advanced benchmark compared with each of the other ethnic groupings (as shown in Table 14). Around one-third of Pasifika students did not reach the low benchmark, while just under a quarter of Mäori students did not reach this low benchmark.

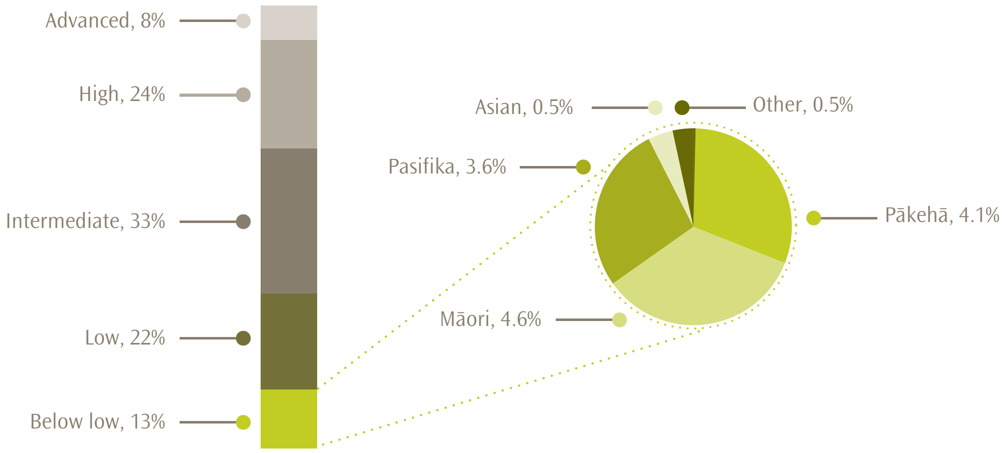
**Table 14** **Proportion of Year 5 students reaching each international benchmark, by ethnic grouping**

| Ethnic grouping | Percentage of Year 5 students reaching each benchmark | | | |
|---|---|---|---|---|
| | Advanced | High | Intermediate | Low |
| Päkehä/European | 10 (0.7) | 41 (1.2) | 76 (1.2) | 93 (0.8) |
| Mäori | 2 (0.7) | 14 (1.9) | 44 (2.7) | 76 (2.6) |
| Pasifika | 1 (0.6) | 7 (1.4) | 31 (3.0) | 64 (3.4) |
| Asian | 14 (2.1) | 43 (3.1) | 72 (2.9) | 92 (2.0) |
| Other | 7 (2.5) | 31 (4.2) | 65 (3.7) | 87 (3.5) |

*Note: Standard errors are presented in parentheses.*

Another way of looking at this information is to examine the composition of the group who did not reach the low benchmark. Thirteen percent of New Zealand students did not reach this benchmark as shown in Figure 10. The majority of these students were Päkehä/European (4.1%) or Mäori (4.6%). However, Mäori and Pasifika students were over-represented in this lower-achieving group compared to their proportions in the population.

**Figure 10 Ethnic composition of the students who did not reach the low benchmark**




*Note: The values presented in the pie chart are proportions of the whole population and therefore add to 13%, the proportion of students in the 'below low' group.*

### Trends in mean science achievement for ethnic groupings

Mäori, Pasifika, and Asian students all demonstrated significant gains in science achievement, on average, between 1994 and 2002. In contrast, Päkehä/European and Other students showed no change over the eight years. However, between 2002 and 2006 the average performances of Mäori and Pasifika students returned to the lower levels of achievement observed in 1994. Lower achievement in 2006 compared with 2002 was also observed for students in the Other ethnic grouping (see Table 15). Asian students in 2006 had the same mean science achievement in 2006 as in 2002 and thus maintained the significant increase observed between 1994 and 2002.

**Table 15 Trends in science achievement 1994 to 2006 by ethnic grouping**

| Ethnic grouping | Mean science achievement 1994 | 1998 | 2002 | 2006 | Change 1994 to 2006 |
|---|---|---|---|---|---|
| Päkehä/European | 534 (3.9) | 541 (4.8) | 532 (3.0) | 528 (2.3) | -6 (4.5) |
| Mäori | 457 (12.0) | 478 (8.0) | 496 (5.2) | 459 (4.9) | 2 (13.0) |
| Pasifika | 441 (14.9) | 436 (13.8) | 496 (5.2) | 431 (5.4) | -10 (15.8) |
| Asian | 493 (16.7) | 517 (10.0) | 529 (4.2) | 529 (6.8) | 36 (18.0) |
| Other | 521 (14.2) | 497 (23.0) | 536 (9.9) | 502 (6.7) | -19 (15.7) |

*Note: Standard errors are presented in parentheses.*

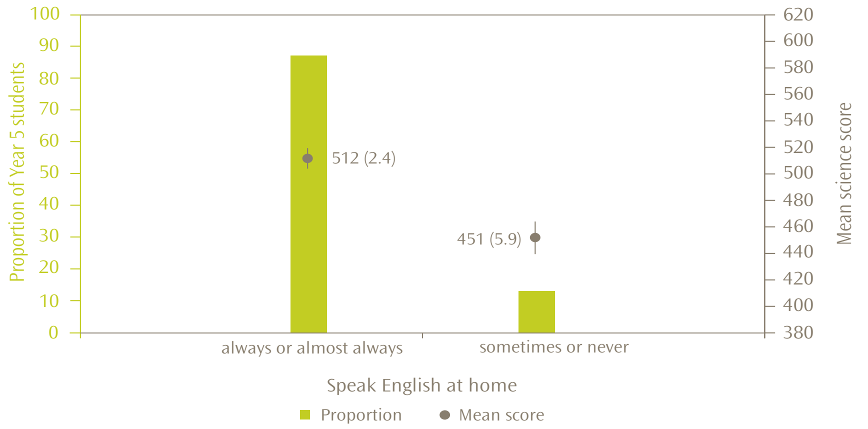
### Science achievement of boys and girls within ethnic groups

As mentioned earlier, there was no significant difference in mean science achievement between boys and girls in TIMSS 2006/07. This result was also observed when gender differences were examined within each of the ethnic groups, with only one exception. For the Päkehä/European, Asian, Mäori, and Other ethnic groupings, there was no significant difference between the girls and the boys. However, Pasifika girls (441) had higher mean science achievement than Pasifika boys (421).

## Science achievement by regularity of English speaking at home

Most students reported that they always or almost always spoke the language of the test (in this case English) at home (87% - 74% always and 13% almost always).[13] Few students (1%) reported that they never spoke English at home. Students who always or almost always spoke English at home had higher science achievement, on average, than those who sometimes or never spoke English at home (see Figure 11). This pattern of higher average achievement for those who spoke English at home was also evident across the previous three cycles of TIMSS (see Caygill, Sturrock, & Chamberlain, 2007). However, it is interesting to note that the difference between these two groups of students has reduced over time from 95 scale score points in 1994 to 61 scale score points in 2006.

**Figure 11 Year 5 mean science scores by regularity of English speaking at home**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. Standard errors are presented in parentheses.*

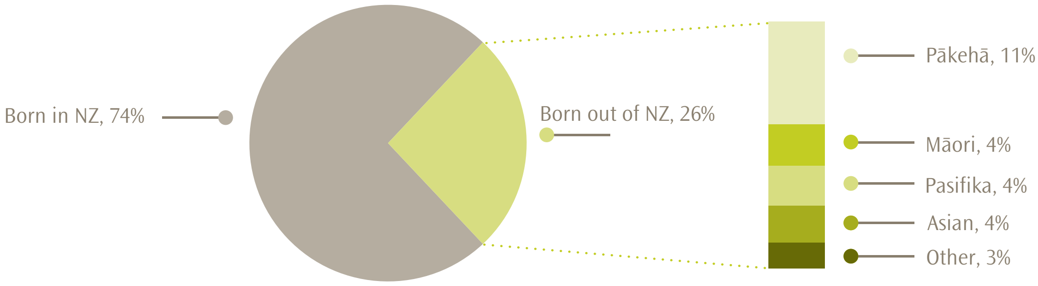
## Science achievement by country of birth

Another factor that interacts with language and ethnicity is the immigrant status of the student and their parents. This information was collected in TIMSS by asking the student if they and their parents were born in New Zealand. Around one-fifth of students had neither parent born in New Zealand, one-fifth had only one parent born in New Zealand and the rest had both parents born in New Zealand. One-quarter of students were born outside of New Zealand. Of these students born out of New Zealand, nearly half of them (44%) came to New Zealand as school-age children.

Science achievement was lower for those students born outside of New Zealand, on average, compared with the New Zealand-born students (46 scale score points difference). The majority of the students born outside of New Zealand were Päkehä/European in ethnic origin, as shown in Figure 12.

13 In TIMSS 2006, as in 1994 and 1998, students who had the majority of their teaching in te reo Mäori were excluded from the assessment. See technical notes and definitions for further details of exclusions.

**Figure 12 Proportions of students born out of New Zealand by ethnic grouping**




*Note: The values presented in the bar are proportions of the whole population and therefore add to 26%, the proportion of students in the 'born out of NZ' group.*

## Interaction of use of English at home, ethnicity and country of birth

In order to confirm the relationships between use of English at home, ethnicity, country of birth and science achievement and also to see how they interacted together, multiple-regression techniques were used. Only these background characteristics were included in the investigation. The resulting statistical model showed that speaking English infrequently at home, belonging to the Mäori or Pasifika ethnic grouping, and being born outside of New Zealand, were all associated with lower science achievement when other factors were taken into account.[14] The model also demonstrates that all of these three background characteristics were significant when explaining differences in science achievement. Note that differences in achievement were smaller when the other factors were taken into account. For example, the difference between those who regularly spoke English at home and those who did not was reduced from 61 scale score points, when analysed in isolation, to 38 scale score points in the model. However, there are a limited number of factors included in this model. Taking into account socio-economic or educational resource factors may change this result (see the section later in this report entitled *Discussion of interactions*).

14 The model showed that when the other factors were taken into account, on average, speaking English infrequently (-38 scale score points or ssp), Mäori (-63 ssp), Pasifika (-72 ssp), born outside New Zealand (-37 ssp) were all associated with lower achievement.

## Science achievement by socio-economic status and home educational resources

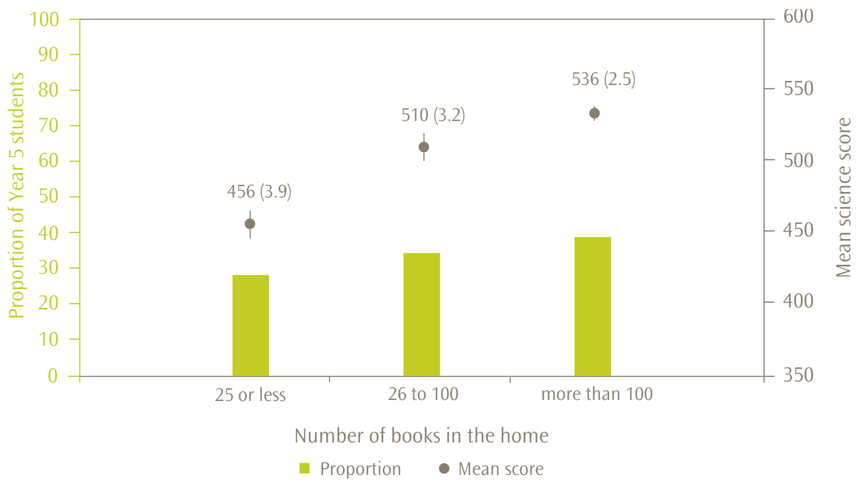
TIMSS includes a number of questions about resources available in the home. These resources in the home can be used as a proxy measure for socio-economic status. In addition, in New Zealand the decile indicator of schools is available to give a measure of the socio-economic status of the area in which a student lives. This section will present analyses of these proxy measures of socio-economic status and their association with science achievement.

### Number of books in the home

Just over one third of New Zealand Year 5 students (38%) reported having more than 100 books in their homes in 2006. This was a large reduction in proportion since 1994 when 62 percent reported having more than 100 books in their homes, but is consistent with 2002 and with that previously found by Caygill and Chamberlain (2005) in the 2001 Trends in Reading Literacy Study (38% also). This trend of fewer books in the home is also consistent with other countries that have been in the study since 1994. Thirty-four percent of New Zealand students reported having between 25 and 100 books in their homes while 28 percent of students reported having 25 or fewer books in their homes.

As shown in Figure 13, there was a positive relationship between the number of books in the home and achievement in 2006, with those students with a greater number of books in the home having higher achievement, on average, in science. This is consistent with findings from previous cycles of TIMSS as well as other studies which have shown a strong link between books in the home and achievement (see for example Chamberlain, 2008).

**Figure 13 Proportions and mean science achievement of Year 5 students by number of books in the home**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. Standard errors are presented in parentheses.*

## Number of items in the home including educational resources

Students were asked whether their home contained items from a list of nine: *calculator, computer (do not include PlayStation®, GameCube®, XBox®, or other TV/video game computers), study desk/table for your use, dictionary, internet connection, your own room, your own mobile phone, musical instruments (e.g., piano, violin, guitar),* and *dishwasher*. The intention of this question was two-fold. The first four items were included to give an indication of the availability of resources at home that could be used to help educationally. The list in its entirety was included to give a proxy measure of socio-economic status as the students were too young to give reliable information on parental employment or household income.

### Items in the home

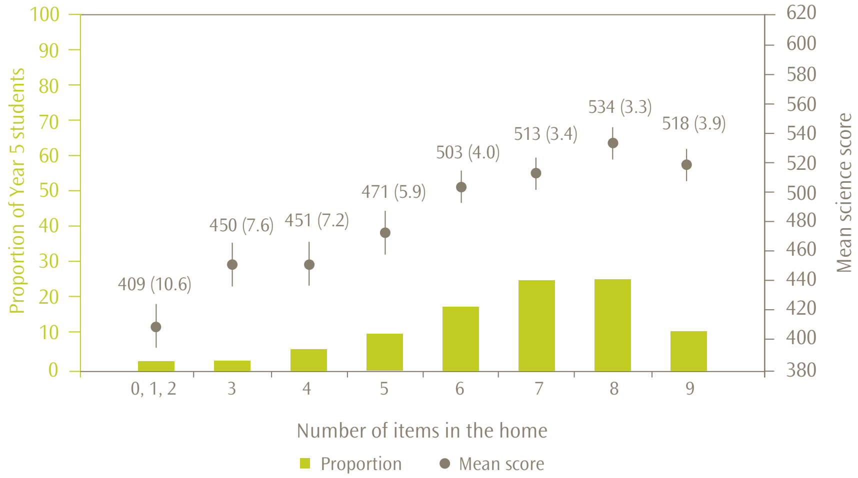
As Table 16 shows, the educational items were the most common items found in the homes of New Zealand Year 5 students. Approximately nine out of every ten students had a calculator in their home and a similar proportion reported having a computer in their home. Just over three-quarters of students reported an internet connection in their home and similarly, three-quarters reported that they had their own room. It was least common for students to have their own mobile phone with only 36 percent of students reporting this.

**Table 16 Proportion of students reporting item is in the home**

| Item | Proportion of Year 5 students (%) |
|---|---|
| Calculator | 92 |
| Computer | 91 |
| Dictionary | 89 |
| Study desk/table for your own use | 80 |
| Internet connection | 77 |
| Your own room | 77 |
| Dishwasher | 69 |
| Musical instruments | 62 |
| Your own mobile phone | 36 |

Eleven percent of students reported that all nine items could be found in their homes, one-quarter reported eight items, and a further quarter reported seven items. Just under 40 percent of students reported six or fewer of the items could be found in their homes, with less than one percent reporting one or none of the listed items. As shown in Figure 14, science achievement generally increased as the number of items in the home increased.

**Figure 14** **Proportions and mean science achievement of Year 5 students by number of items found in the home**



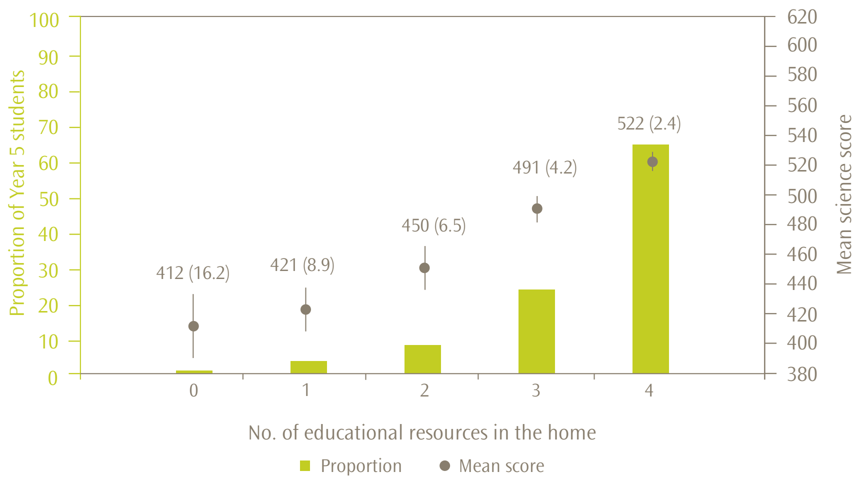

Note: *The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies.*
*Standard errors are presented in parentheses.*

### Home educational resources

As mentioned earlier, the *calculator, computer (do not include PlayStation®, GameCube®, XBox®, or other TV/ video game computers),* and *study desk/table for your use* were included in the list of items to ascertain the availability of educational resources at home. While students may not necessarily use these items for educational purposes, the presence of these items could indicate the relative importance of education to the family, although this may also be reflective of the wealth of the home.

Nearly two-thirds of students (64%) reported that they had all four educational resources in their homes, while nearly one-quarter of students (24%) reported three of the four items in their homes. As shown in Figure 15, students with more educational resources in the home had higher mean science achievement than those with fewer items.

**Figure 15** **Proportions and mean science achievement of Year 5 students by number of educational resources found in the home**



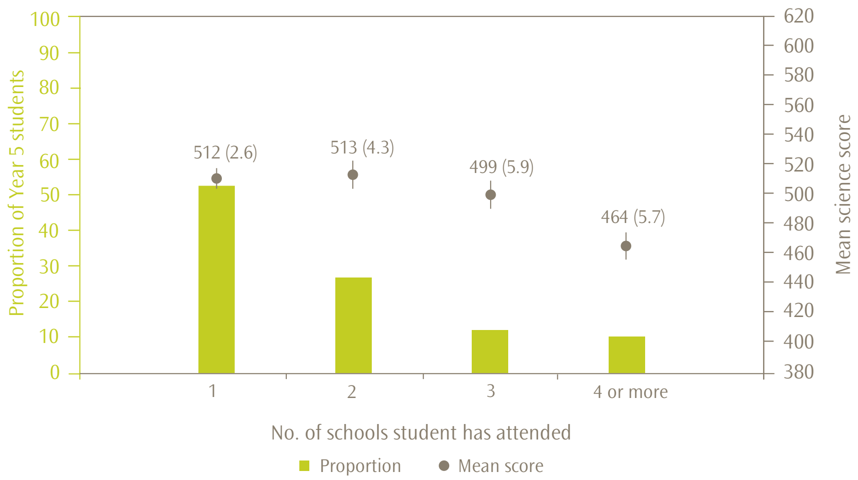

Note: *The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. The achievement value for the students with 0 educational resources should be treated with caution as indicated by the high standard error around the mean.*
*Standard errors are presented in parentheses.*

## Number of schools attended by student

Another question included in the questionnaire that may be indicative of socio-economic status was the number of schools attended by a student. Many students in New Zealand change schools for a variety of reasons, but high mobility may be symptomatic of families moving regularly to find work, or students moving about among family members, or in care.

Just over half of all students reported that they had only attended one school, their current school, while one quarter of students had attended two schools. One in every 10 students reported they had attended four or more schools. As shown in Figure 16, students with high mobility had lower achievement than those with low mobility.

**Figure 16 Proportions and mean science achievement of Year 5 students by number of schools student has attended**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. Standard errors are presented in parentheses.*

## Household size

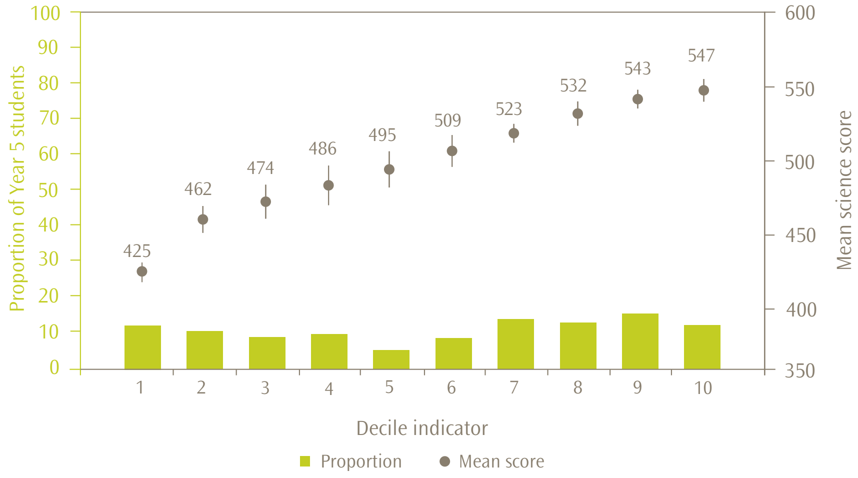
Another indicator of socio-economic status is the size of the household. While cultural or religious beliefs may determine household size, household crowding may also be indicative of poorer economic background. However, homes with one child living with one parent may also struggle financially. Science achievement was examined with respect to number of people in the household. The highest achievement was found amongst students in households of size 3, 4, or 5 (513, 525, and 517 scale score points respectively) with the lowest achievement amongst students in households of size 2, 7 or 8 (477, 472 and 439 respectively).

## Decile

The Ministry of Education allocates resources such as Targeted Funding for Educational Achievement (TFEA) based on school decile indicator. A school's decile indicates the extent to which a school draws its students from low socio-economic communities. In general, decile 1 schools are the schools with the highest proportion of students from socio-economically disadvantaged communities, while decile 10 schools are the ten percent of schools with the lowest proportion of students from these communities.

Analyses of science achievement for students in schools in each decile band demonstrate that science achievement is higher in higher decile schools and lower in lower decile schools as shown in Figure 17. However, the difference in achievement was not always significant when adjacent groups were examined. The largest difference between the mean science score of adjacent groups occurs between students in deciles 1 and 2 (37 scale score points). It should be noted that this analysis does not demonstrate a causal link between being in a higher decile school and having higher achievement. Rather it is indicative of a trend demonstrating that students with lower levels of disadvantage in terms of family background and socio-economic background and living in wealthier areas have higher achievement.

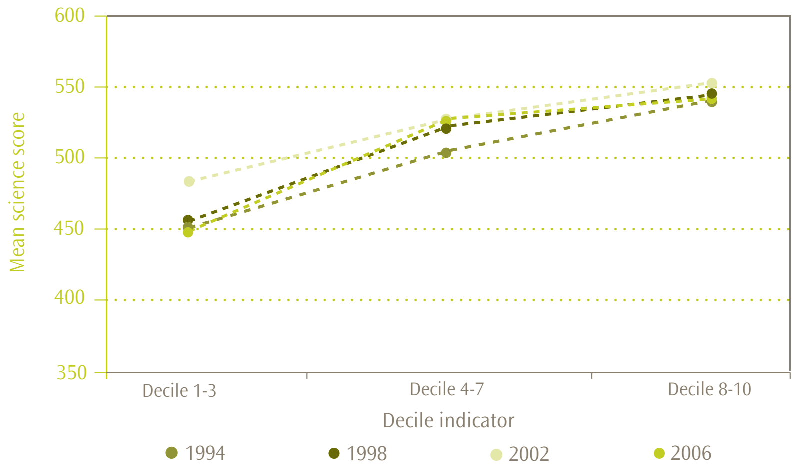
**Figure 17** **Proportions and mean science achievement of Year 5 students by decile indicator of school**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies.*

Decile results were not reported in earlier TIMSS studies at this level. However, they have been analysed for this report by the broad groupings 1 to 3, 4 to 7, and 8 to 10 in order to ascertain whether the pattern of higher achievement, on average, among students in higher decile schools, was also evident in earlier cycles. As shown in Figure 18, students in higher decile schools have consistently demonstrated higher achievement than those in lower decile schools.

**Figure 18** **Trends in science achievement across decile groups for 1994 to 2006**




## Summary of science achievement by socio-economic status and home educational resources

As the results in this section demonstrate, students from higher socio-economic backgrounds tend to have higher mean science achievement than those from lower socio-economic backgrounds as evidenced by the proxy measures *books in the home, items in the home, household size* and *mobility*. In addition, the decile of the school they attend, indicative of the level of economic disadvantage in the community in which they live, was positively related to science achievement. That is students in higher decile schools had higher science achievement, on average, than those in lower deciles.

## Student activities outside of school

Previous cycles of TIMSS have shown that watching television and videos was the most popular leisure activity for Year 5 students (see Caygill, Sturrock, & Chamberlain, 2007). Leisure activities are of interest in TIMSS as they can provide positive learning experiences as well as the negative implications of reducing time for doing school-related learning at home. While no judgements are made in this section of the value of leisure activities, and acknowledging that learning occurs within and outside of school, it is interesting to look at the changes over time and also to examine the relationships with achievement.

Table 17 presents the mean number of hours per school day that students reported spending on a variety of activities, along with the proportion of students who reported spending more than 2 hours on each activity. Note that it is possible that some of these activities were 'multi-tasked'. For example, students who spent an hour after school playing sports with friends selected *playing sports* and *playing or talking with friends*. Playing or talking with friends and playing sports were the two most popular activities for Year 5 students in 2006. Television watching was relegated to third most popular activity in 2006 in contrast with previous cycles of TIMSS.

**Table 17** **The proportion and mean amounts of time Year 5 students reported spending on leisure activities**

| Leisure activities | Mean number of hours per school day | Proportion spending more than 2 hours (%) |
|---|---|---|
| Watching television and videos | 1.5 (0.03) | 25 |
| Playing computer games | 1.0 (0.02) | 15 |
| Playing or talking with friends | 1.7 (0.03) | 31 |
| Doing jobs at home | 1.3 (0.03) | 20 |
| Playing sports | 1.7 (0.03) | 30 |
| Reading a book for enjoyment | 1.2 (0.02) | 18 |
| Using the internet | 1.1 (0.02) | 16 |

*Note: Mean number of hours based on: No time=0; Less than 1 hour=0.5; 1-2 hours=1.5; More than 2 but less than 4 hours=3; 4 or more hours=4.5. Activities are not necessarily exclusive.*
*Standard errors are presented in parentheses.*

The relationship between the number of hours spent in the individual activities and science achievement was relatively consistent across the activities. Students who reported a small or moderate amount of time in an activity generally had higher achievement than those who reported no time or many hours on the activity.

## Student attitudes

Students were asked how much they agree with eight statements about learning science (listed in Table 18 – positive and negative statements were interwoven in the questionnaire but are reordered here for easier reading). They were given four response options: *agree a lot, agree a little, disagree a little, disagree a lot.*

Students were generally positive about science with 84 percent agreeing that they enjoy learning science and 78 percent disagreeing that science was boring. The majority of students agreed that they would like to do more science in school (81%). This is similar to the National Educational Monitoring Project (NEMP) findings at the Year 4 level, where 71 percent of students indicated that they would like to do more science (Crooks, Smith & Flockton, 2008).[15] In contrast with this positive response, around four in every ten students agreed that science was harder for them than for many of their classmates.

**Table 18 Proportion of students who responded positively to statements about learning science**

| Statements about learning science | Proportion of students | |
|---|---|---|
| Positive statements | Agreeing (%) | Disagreeing (%) |
| I usually do well in science | 75 (0.9) | 25 (0.9) |
| I would like to do more science in school | 81 (0.7) | 19 (0.7) |
| I enjoy learning science | 84 (0.7) | 16 (0.7) |
| I learn things quickly in science | 70 (0.7) | 30 (0.7) |
| I like science | 82 (0.7) | 18 (0.7) |
| Negative statements | Agreeing (%) | Disagreeing (%) |
| Science is harder for me than for many of my classmates | 41 (1.0) | 59 (1.0) |
| I am just not good at science | 32 (0.9) | 68 (0.9) |
| Science is boring | 22 (0.7) | 78 (0.7) |

Note: *The values for agree combine student responses to 'agree a lot' and 'agree a little'. Similarly the values for disagree combine 'little' and 'a lot'.*
*Standard errors are presented in parentheses.*

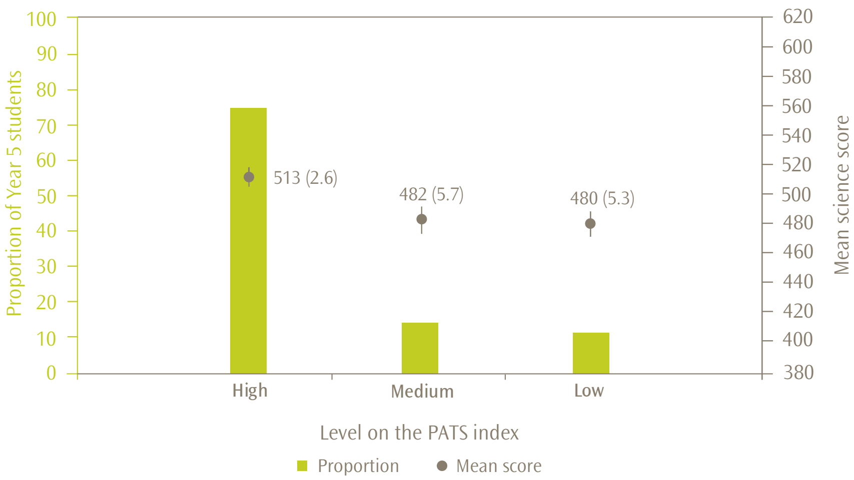
Generally, students with positive attitudes towards science had higher achievement than students with negative attitudes. In order to examine the relationship with achievement, the international researchers combined the data in two indices: the index of students' positive affect toward science and the index of students' self-confidence in learning science.

15 Students were asked in NEMP "Would you like to do more or less science at school" with response options *more, about the same,* and *less.*

## Index of students' positive affect toward science

The three statements: I enjoy learning science; science is boring; and I like science; were combined to form the index of students' positive affect toward science (PATS).[16] Three-quarters of the students were at the high level of this index; that is, on average, they were positive about science. Eleven percent of students were at the low level of the index; that is, on average, they were negative about science. These proportions have not changed since 1994. As shown in Figure 19 students who were more positive about science (at the high level of the PATS index) had higher mean science achievement than those that were more negative. There was no difference between those at the medium and low levels of the index.

**Figure 19 Proportion and mean science achievement of students at each level of the positive affect toward science (PATS) index**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. Standard errors are presented in parentheses.*

In relation to other countries, relatively low proportions of New Zealand students were at the high level of the PATS index. That is, few students reported positive attitudes towards science in comparison to their international counterparts. However, this proportion was the same as the United States, Chinese Taipei, Austria, and Singapore (all 75%). England had a lower proportion of students who responded positively to these statements (59% at the high level of the PATS index).

## Index of students' self-confidence in learning science

The four statements: I usually do well in science; science is harder for me than for many of my classmates; I am just not good at science; and I learn things quickly in science; were combined to form the index of students' self-confidence in learning science (SCS).[17] Just over half (51%) of the students were at the high level of this index; that is, on average, they were positive about their own abilities in science. Twelve percent of students were at the low level of the index; that is, on average, they were negative about their abilities in science.

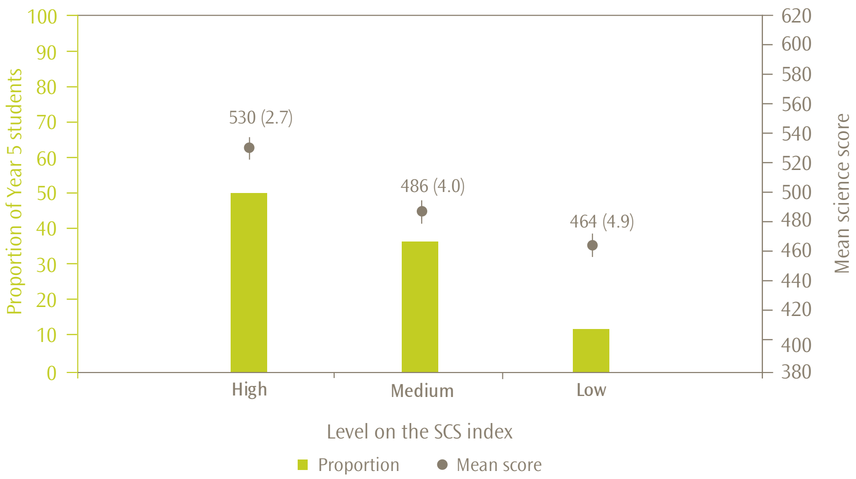
The proportions of the students at all levels of the index have changed significantly since 2002. More students are now positive about their abilities to learn science (15 percentage point increase), but also more students are negative about their abilities to learn science (7 percentage point increase). Fewer students are, therefore, at the medium level of the index.

16 An average was computed across a 4-point scale with 1 agree a lot, 2 agree a little, 3 disagree a little, 4 agree a lot. The statement 'science is boring' was reversed so that students disagreeing a lot were given a value of 1.

17 An average was computed across a 4-point scale with 1 agree a lot, 2 agree a little, 3 disagree a little, 4 agree a lot. The statements 'science is harder for me than for many of my classmates' and 'I am just not good at science' were reversed so that students disagreeing a lot were given a value of 1.

As shown in Figure 20, students who were more positive about their abilities to learn science (at the high level of the SCS index) had higher mean science achievement than those that were more negative. Those students with the lowest self-confidence had the lowest achievement on average. Note that the difference in mean science achievement between students that were high and those that were low on the SCS index (66 scale score points) is greater than those in the respective groups on the PATS index (33 scale score points). Thus the self-confidence of students had a stronger relationship with science achievement than having a positive attitude towards science.

**Figure 20 Proportion and mean science achievement of students at each level of the students' self-confidence in learning science (SCS) index**




*Note: The bars on the graph represent the proportions of Year 5 students while the points represent mean scores. Lines extending from the points represent the 95% confidence interval, i.e. the range within which we are 95 percent confident that the true population value lies. Standard errors are presented in parentheses.*

In relation to other countries, very low proportions of New Zealand students were at the high level of the SCS index. However this proportion was similar to those for Hong Kong SAR (52%) and Japan (53%). In comparison, the United States had a much larger proportion of students who responded positively to these statements about their abilities in science (69% at the high level of the SCS index).

## Attitudes to science by gender

Boys and girls demonstrated very similar attitudes to science. Around three-quarters of girls and boys were very positive about science and were at the high level of the PATS index (76% and 75% respectively). Around half of girls and boys reported confidence in their science abilities and were at the high level of the SCS index (50% and 51% respectively).

## Attitudes to science by ethnicity

Some differences were evident among the ethnic groupings when attitudes to science were considered. More students in the Other ethnic grouping reported positive attitudes to science and were at the high level of the PATS index (82% compared with 76% of Päkehä/European, 75% of Asian, 74% of Pasifika, and 73% of Mäori). There were no significant differences in the proportion of students at the high level of the index when Päkehä/European, Asian, Pasifika, and Mäori students were compared.

More Päkehä/European students and students in the Other ethnic grouping reported higher levels of self-confidence and were at the high level of the SCS index (55% and 53% respectively) compared with their Asian, Mäori, and Pasifika counterparts (46%, 44%, and 42% respectively). There were no significant differences in the proportion of students at the high level of the index when Asian, Pasifika, and Mäori students were compared.

## Discussion of interactions

This report so far has presented results related to the science achievement of New Zealand Year 5 students from TIMSS 2006/07, mainly in the form of descriptive statistics. The focus has been on background characteristics of the students in isolation from other characteristics, when in fact, many of these characteristics are interrelated. For example, one could speculate that differences in achievement between students of different ethnic groups are interrelated with differences in home language. Indeed, earlier in this report it was demonstrated by analysing these factors together, that the difference in science achievement between those who spoke English at home regularly and those who did not was smaller when ethnic differences and immigration status were taken into account.

To investigate the possible interactions between characteristics of students and science achievement outcomes further, some statistical modelling was undertaken. This involved putting all of the factors that have shown a relationship with achievement together in a statistical model using an analysis tool called MLWin.[18] This tool carries out multi-level modelling analysis of this type of data. A more detailed discussion of the modelling work will be presented in a separate working paper later in 2008 or early in 2009.

The value of looking at such models is that the relative importance of different background characteristics can be determined. Modelling also allows for the elimination of factors that are unimportant or measure the same underlying trait as others. The multi-level aspect of this type of statistical model takes account of the fact that student learning takes place within classes that are part of schools and that all of these can impact on achievement. Thus this type of model allows for consideration of differences in results between schools, classes, and students.

The first step of the analysis was to examine variations in science achievement between schools, classes, and students, without including any background characteristics. Around 22 percent of the variation in science achievement was attributable to differences between schools, around 7 percent to differences between classes in the same school, and around 70 percent to differences between students. In other words, while there were some differences in science achievement between the schools, and also the classes within schools, the majority of differences were between the students themselves.

In the second step of the analysis, the following background characteristics were considered in the model: gender, ethnicity, speaking English at home, born outside of New Zealand, student age, books in the home, possessions in the home, attitudes, out-of-school activities, school decile and science instructional hours.

The model of influences on Year 5 students' science achievement shows the following significant relationships with achievement when the other factors were taken account of:

- students with a greater number of books in the home had higher science achievement than those who had fewer;
- older students had higher science achievement than younger students;

18 For further details see Goldstein (2003).

- students born in New Zealand had higher science achievement than those who were not;
- students with a greater number of educational resources in the home had higher science achievement than those with fewer;
- students who spoke English frequently at home had higher science achievement than those who did not;
- students with high self-confidence had higher science achievement than those with lower self-confidence;
- students who felt safe at school had higher science achievement than those who felt less safe;
- students who engaged in the out-of-school activities television watching and playing computer games for a moderate amount of time had higher science achievement than those who never did or did these for a higher number of hours per day;
- students who read books regularly had higher science achievement than those who did not;
- boys had higher science achievement than girls;
- Asian students had higher science achievement than Päkehä/European students;
- Mäori students had lower science achievement than Päkehä/European students;
- Pasifika students had lower science achievement than Päkehä/European students;
- students who were in a higher decile school (living in a community with less economic disadvantage) had higher science achievement than those in lower decile schools.

After completing the statistical model, variations between schools, classes, and students were re-examined to see if the model helped to explain the variations initially observed. The model explained just under three-quarters of the variation between schools, just under one-half of the variation between classes, and about one-quarter of the variation between students. In other words, the model explained much of the variation between schools, and between classes within schools, so that most of the variation that remained unexplained was between students.

Factors relating to the socio-economic status of the students such as decile, educational resources in the home, and books in the home all had a reasonably strong relationship with science achievement in the model. While it is not easy to effect change in these, self-confidence in science and reading books were two factors that had a strong relationship with achievement that might be influenced by teachers and parents.

This report has not examined all of the data collected in TIMSS. Further analyses will be undertaken including investigation of what is happening in science in schools from information collected from teachers and principals. The model presented in this section is an initial investigation and should be read as such, but it gives valuable insight into the factors explaining differences in the science achievement of Year 5 students in New Zealand.

# Conclusion

This report has examined trends in New Zealand science achievement at the Year 5 level from 1994 to 2006. It has looked at New Zealand Year 5 students' science achievement in relation to other countries that participated in the study. An examination of the TIMSS assessment questions in relation to New Zealand's science curriculum was presented along with analyses of achievement by sub-groupings (such as gender and ethnicity) and background factors. A statistical model that attempts to explain variations among students, classes, and schools using the background information discussed in this report was also described.

## Achievement in science

Overall, the mean science achievement of New Zealand Year 5 students was about the same in 2006 as in 1994. Although an increase in mean science achievement was observed in 2002 relative to earlier years, this performance was not sustained in 2006. In terms of the distribution of science achievement across the range of scores, this was narrower in 2006 than in 1994. A positive aspect of this change is that fewer students are demonstrating very low achievement, but it also means a smaller proportion of New Zealand students are gaining very high scores. In international terms, New Zealand Year 5 science achievement is significantly higher than 13 of the 36 countries participating in TIMSS at the middle primary level, but significantly lower than 21 of the 36 countries.

Year 5 students continue to demonstrate relative strengths in aspects of science. They tend to perform relatively better on *earth science* questions compared to *life* and *physical science*. Students also perform relatively better on questions that involve *demonstrating* knowledge compared to questions that assess *applying* or *reasoning*.

## Science instructional hours

There has been a change in the number of hours that teachers reported spending on science instruction with New Zealand Year 5 students since 2002. In 2006, the number of hours was significantly lower than in 2002. This reduction in the number of hours that students are engaged in science in the middle primary years is consistent with evidence from other studies that report on students' experiences of science in the classroom (e.g. NEMP).

## Background characteristics

Both high and low performers were found among boys and girls, and in all ethnic groupings. On average, there was no difference in science achievement between boys and girls. However, some differences were observed among the ethnic groupings. Päkehä/European and Asian students had similar science achievement, on average, and their mean science achievement was higher, on average, than that of Mäori and Pasifika students. Mäori students had higher mean science achievement than Pasifika students.

In terms of other background characteristics, science achievement was higher, on average, among students who regularly spoke English at home. Similarly, students who were born in New Zealand had higher science achievement, on average, than those who were not. Students from higher socio-economic backgrounds tended to have higher mean science achievement than those from lower socio-economic backgrounds as evidenced by the proxy measures *books in the home*, *items in the home*, *household size* and *mobility*. In addition, the decile of the school they attended, indicative of the level of economic disadvantage in the community in which they live, was positively related to science achievement. That is, students in higher decile schools had higher achievement, on average, than students in lower decile schools. Year 5 students who reported a small or moderate amount
of time in out-of-school leisure activities generally had higher achievement than those who reported no time or many hours on the activity.

## Attitudes to science

New Zealand Year 5 students generally expressed positive attitudes towards science. Consistent with NEMP (Crooks, Smith, & Flockton, 2008), eight out of every ten students indicated that they would like to do more science in school. Those students who reported positive attitudes towards science or were confident in their own science abilities had higher achievement than those who were less positive or confident.

Boys and girls expressed similar attitudes to science, both in terms of enjoyment and motivation, and self-confidence. More Päkehä/European and students in the Other ethnic grouping reported high self-confidence in science compared with Asian, Mäori, and Pasifika students. More students in the Other ethnic grouping reported positive attitudes towards science compared with Päkehä/European, Asian, Mäori, and Pasifika students.

## Final comment

The Ministry of Education's current focus is on presence, engagement and achievement (Ministry of Education, 2007). The Ministry has in place a number of mechanisms for monitoring student performance in science in primary and secondary schooling at the system level. Performance in science reflects learning from within school, in family and whänau, and in the broader community. While overall there has been little change in the data on student achievement, the reduction in science teaching hours along with other changes and variations in achievement among groupings requires further attention.

This report has not examined all of the data collected in TIMSS. Further analyses will be undertaken including investigation of what is happening in science in schools from information collected from teachers and principals. Further reports will become available during 2009.

## References


Caygill, R., & Chamberlain, M. (2005). *Reading achievement in New Zealand in 1990 and 2001: Results from the Trends in IEA's Reading Literacy Study*. Wellington: Ministry of Education.

Caygill, R., & Kirkham, S.A. (2008). *Trends in Year 5 mathematics achievement 1994 to 2006: results from the Trends in International Mathematics and Science Study (TIMSS)*. Wellington: Ministry of Education.

Caygill, R., Sturrock, F., & Chamberlain, M. (2007). *Mathematics and science achievement in New Zealand: tracking the changes of Year 5 students in TIMSS 1994-2002*. Wellington: Ministry of Education.

Chamberlain, M. (2008). *PIRLS 2005/2006 in New Zealand: an overview of national findings from the second cycle of the Progress in International Reading Literacy Study*. Wellington: Ministry of Education.

Geske, A., Grinfelds, A., & Ozola, A. (2008). Latvia. In I.V.S. Mullis, M.O. Martin, J.F. Olson, D.R. Berger, D. Milne, & G.M. Stanco (Eds.), *TIMSS 2007 encyclopedia: a guide to mathematics and science education around the world* (pp. 347-354). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Goldstein, H. (2003). *Multilevel Statistical Models*. 3rd ed. London: Arnold.

Japelj Pavešić, B., & Svetlik, K. (2008). Slovenia. In I.V.S. Mullis, M.O. Martin, J.F. Olson, D.R. Berger, D. Milne, & G.M. Stanco (Eds.), *TIMSS 2007 encyclopedia: a guide to mathematics and science education around the world* (pp. 561-572). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., & Foy, P. (2008). *TIMSS 2007 international science report: findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Ministry of Education. (1993). *Science in the New Zealand curriculum*. Wellington: Learning Media Ltd.

Ministry of Education. (2007). *Statement of intent 2007-2012*. Wellington: Ministry of Education.

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Quek, G., Goh, K.H., Yen, Y. P., Liu, Y.M., Tan, Q.L., Chin, T.Y., & Lau, C.Y. (2008). Singapore. In I.V.S. Mullis, M.O. Martin, J.F. Olson, D.R. Berger, D. Milne, & G.M. Stanco (Eds.), *TIMSS 2007 encyclopedia: a guide to mathematics and science education around the world* (pp. 537-560). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

United Nations Development Programme. (2008). *Human development report 2007/2008*. (pp. 229-232).

# Definitions and technical notes

This section gives a brief overview of the technical details and definitions applicable to this report. For a comprehensive description of the technical details pertaining to TIMSS see the *TIMSS 2007 technical report* (Olson, Martin, & Mullis, (Eds.), 2008)

## Benchmarks

In order to describe more fully what achievement on the science scale means, the TIMSS international researchers have developed benchmarks. These benchmarks link student performance on the TIMSS science scale to performance on science questions and describe what students can typically do at set points on the science achievement scale. The international science benchmarks are four points on the science scale, the advanced benchmark (625), the high benchmark (550), the intermediate benchmark (475), and the low benchmark (400). The performance of students reaching each benchmark is described in relation to the types of questions they answered correctly.

## Exclusions

Each country was permitted to exclude some students for whom the assessment was not appropriate or was difficult to administer. Countries were required to keep the amount of excluded students as small as possible, with a guideline of 5 percent of the 'target' population as the maximum. Any countries that exceeded this value are indicated in the international exhibits. The target population in New Zealand was Year 5 students.

School-level exclusions in New Zealand consisted of very small schools (less than 4 Year 5 students), special education schools, Rudolf Steiner schools, the Correspondence School, and schools that provide more than 80% of their instruction in te reo Mäori. Within-school exclusions consisted of special education classes, special needs students, students with insufficient instruction in English, and units within schools that provide more than 80% of their instruction in te reo Mäori.

The New Zealand exclusion rate was one of the largest at 5.4 percent and equivalent to Hong Kong SAR and Lithuania. Exclusion rates for most of the other countries were usually kept below the 5 percent maximum, with only the United States and the benchmarking participants exceeding this level.[19]

## Making models

The models in this report were formulated using two different methods. Regression analyses were used for the model at the student level that combined ethnicity, speaking English at home, and immigrant status. Custom-written programs described in the *TIMSS user guide for the international database* (to be published in early 2009) were used for this analysis. Multi-level modelling techniques were applied using the MLWin package for the analysis which examined school-, class-, and student-level variations in achievement. A range of background characteristics were included in the larger model initially and the model was then tested iteratively. At each iteration, any characteristics that were not statistically significant were removed until the model contained only variables with a significant influence on student achievement.

19 See Martin, Mullis, & Foy, P. (2008), Exhibit A.4 for this information.

### Mean, medians, and averages

There are three possible measures of central tendency, but only the mean and the median are used in this report.

The mean of a set of scores is the sum of the scores divided by the number of scores, and is also sometimes referred to as 'the average', particularly in the international reports. Note that for TIMSS, as with other large-scale studies, the means for a country are adjusted slightly (in technical terms 'weighted') to reflect the total population of Year 5 rather than just the sample.

A median is the middle number when all numbers are put in order.

In earlier cycles of TIMSS, an international mean was reported. However, as the number of countries participating changed, this mean shifted so that it was difficult to make comparisons across years. In TIMSS 2006/07 the TIMSS scale average is reported. This is the value to which the scores of each student are scaled (see later note on *Scale score points* for more details).

### Minimum group size for reporting achievement data

In this report, student achievement data are not reported where the group size is less than 30 students or less than 10 schools. While group sizes of 30 to 50 students do have achievement reported in some cases, these are annotated and should be treated with caution as there is a lot of uncertainty in the measurement, as demonstrated by larger standard errors.

### Percentile

The percentages of students performing below or above particular points on the scale can be used to describe the range of achievement. The lowest outer limit of achievement reported in ranges is the 5th percentile – the score at which only 5 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 5 percent of students achieved a higher score and 95 percent of students a lower score. Therefore 90 percent of the Year 5 student scores lie between the 5th and 95th percentiles.

### Sampling

Schools are sampled in TIMSS with a probability proportional to the number of Year 5 students. In order to improve the precision of sampling, the schools were ordered by decile, level of urbanisation, and size, so that the schools selected better represented the population of schools in New Zealand. Within each school, classes were sampled with equal probability and all Year 5 students within each class were selected.

### Scale score points

The design of TIMSS allows for a large number of questions to be used in mathematics and science; each student answers only a portion of these questions. TIMSS employs techniques to enable population estimates of achievement to be produced for each country even though a sample of students responded to differing selections of questions. These techniques result in scaled scores that are on a scale with a mean of 500 and a standard deviation of 100.

## Significance tests

In this report, all the comparisons that have been made are tested for statistical significance using the *t* statistic, with the probability of making an incorrect inference set at 5 percent. To compare the means of two groups of students, the formula to generate the test statistics computed in this report is:

(1) $$t = \frac{\bar{x}_1 - \bar{x}_2}{se_{diff}}$$

The calculation of $se_{diff}$, the standard error of the difference, varies depending on whether the groups were sampled independently or not. If the means for two groups that were sampled independently are being compared, for example, boys' achievement in 1994 and 2006, then the standard error of the difference is calculated as the square root of the sum of the squared standard errors of each mean:

(2) $$se_{diff} = \sqrt{se_1^2 + se_2^2}$$

For most of the comparisons, this formula was not applicable and so the sediff is computed more accurately by combining variances using custom-written SAS programs. However as a rough estimate, the above formula will give a similar result.

Note that in all calculations, unrounded figures are used in these tests, which may account for some results appearing to be inconsistent.

## Standard error

Because of the technical nature of TIMSS, the calculation of statistics such as means and proportions has some uncertainty due to:

(i) generalising from the sample to the total Year 5 school population; and
(ii) inferring each student's proficiency from their performance on a subset of questions.

The standard errors provide a measure of this uncertainty. In general, we can be 95 percent confident that the true population value lies within an interval of 1.96 standard errors either side of the given statistic. This confidence interval is represented in graphs by the lines extending in either direction from the points.

## Statistically significant

In order to determine whether a difference between two means is actual, it is usual to undertake tests of significance. These tests take into account the means and the error associated with them. If a result is reported as not being statistically significant, then, although the means might be slightly different, we do not have sufficient evidence to infer that they are different. All tests of statistical significance referred to in this report are at the 95 percent confidence level.

## Weighting

Due to the use of sampling, weights need to be applied when analysing the TIMSS data. Weighting ensures that any information presented more closely reflects the total population of Year 5 students rather than just the sample. The TIMSS weighting takes into account school, class, and student level information and the overall sampling weight is a product of the school, class, and student weights.

First published in December 2008 by:
Comparative Education Research Unit
Research Division
Ministry of Education
PO Box 1666
Wellington
New Zealand

Email: research.info@minedu.govt.nz
Fax: 64-4-463 8312
Phone: 64-4-463 8000



This report is available from the Education Counts website:
www.educationcounts.govt.nz

ISBN 978-0-478-13968-6
ISBN 978-0-478-13969-3 (web)
RMR-914