CHAPTER 1: Introduction

KEY POINTS

- "TIMSS-98/99", also known as "TIMSS-R", was a partial replication of the earlier Third International Mathematics and Science Study (TIMSS-94/95).
- TIMSS-98/99 was an international study designed to measure trends in student achievement at the lower secondary level — in New Zealand's case, it involved Year 9 students.
- The mathematics and science teachers of the participating Year 9 students, as well as the principals of the schools involved, also provided information.
- Thirty-eight countries took part in TIMSS-98/99; 26 of these countries also took part in TIMSS-94/95.
- New Zealand also administered a national version of "TIMSS-R" at the middle primary school level, involving mostly Year 5 students — the results from that study are the subject of a separate report (forthcoming).

WHAT IS "TIMSS-98/99"?

"TIMSS-98/99", also known as "TIMSS-R", is a partial replication of the Third International Mathematics and Science Study (TIMSS) — in which New Zealand was involved from 1992 through 1997. TIMSS was a large-scale comparative study involving more than 40 countries. Primarily, it examined student achievement in mathematics and science and students' beliefs about and attitudes towards these subjects. It also investigated curricular intentions, and school and classroom environments. New Zealand participated in virtually all aspects of TIMSS including the curricula study, and the student assessments including a (hands-on) performance assessment study. The study involved students at five class levels — Years 4 and 5 students[1]; Years 8 and 9 students[2]; and senior students in their final year of schooling[3].

WHAT WAS THE MAIN AIM OF TIMSS-98/99?

The main aim of TIMSS-98/99 was to provide comprehensive *trend* information on student achievement in mathematics and science. Whereas TIMSS (or TIMSS-94/95)[4] investigated student achievement at five class levels, TIMSS-98/99 had a narrower focus — that is, it focused just on the class level where New Zealand Year 9 students are located — or using international nomenclature — '*grade 8*'[5]. At its inception, it was envisaged that the TIMSS-94/95 study would be the first of a series of mathematics and science assessments to be conducted every four years. TIMSS-98/99 is the second in the series and the next study is planned for 2002/2003.

TIMSS-98/99 was administered in the Southern Hemisphere and the Northern Hemisphere in adjacent calendar years, and four years on from the respective administration of TIMSS-94/95 at the middle primary level in each country. Therefore, TIMSS-98/99 was, in part, a cohort longitudinal study as it gave some countries the opportunity to examine the progress the younger cohort had made in mathematics and science over the four years. It is, however, important to remember that the students who participated in TIMSS-98/99 are not necessarily the same individuals who took part in TIMSS-94/95.

As well as trends in achievement, TIMSS-98/99 aimed to provide trend information in terms of the context of students' achievement. For example, to what extent had teaching and learning practices changed in four years at the lower secondary level? What had changed in school environments? What was the impact of any curricula reforms during the four years?

What did participation in TIMSS-98/99 mean for New Zealand?

New Zealand's participation in TIMSS-98/99 provided the opportunity to examine Year 9 students' mathematics and science achievement in 1998 relative to that of their Year 9 student counterparts in 1994. As well, since New Zealand Year 5 students participated in TIMSS in 1994, it meant that we could examine the progress this cohort had made in four years, as they were now the 1998 Year 9 cohort.

---

[1] *Using the class level nomenclature the students were in standards 2 and 3; Years 4 and 5 students are mostly in these two class levels.*

[2] *Forms 2 and 3 using the class level nomenclature.*

[3] *These students were in forms 6 and 7.*

[4] *The assessment phase of TIMSS was administered in Southern Hemisphere countries in 1994 and in Northern Hemisphere countries in 1995. TIMSS-94/95 and TIMSS-98/99 will be used to distinguish between the two assessments.*

[5] *New Zealand also chose to replicate TIMSS with Year 5 students (in standard 3) or grade 4 equivalent. The results of this study are the subject of a forthcoming report.*

TO WHAT EXTENT IS TIMSS-98/99 A REPLICATION OF TIMSS-94/95?

The curriculum frameworks, on which the TIMSS-94/95 achievement test was based, remained the same, and were used to develop new test items for TIMSS-98/99 to replace those released into the public domain at the conclusion of TIMSS-94/95. Similarly, the design of the achievement test, the content of questionnaires (with a few exceptions), the sampling design, test administration and data collection, and quality assurance procedures were all replicated in TIMSS-98/99. While some of the details are covered towards the end of this chapter, more information can be obtained from the international technical volumes listed in the references section at the end of this report. Alternatively, information is available on www.timss.org.

THE RELATIONSHIP BETWEEN THE CONTEXT OF TIMSS-94/95 AND THE ADMINISTRATION OF TIMSS-98/99 IN NEW ZEALAND

TIMSS-94/95 assessments that involved primary school and lower secondary school students were administered in New Zealand in late 1994. This was during a period of major curricula reform — the new mathematics curriculum was introduced into schools in early 1993 and schools were still working towards implementation of the new science curriculum[6]. It was generally felt, however, that the new curricula would have had very little impact on the achievement results for New Zealand students (eg, Garden 1996a & 1996b). Six reports containing national results from TIMSS-94/95 for New Zealand were published, over the period 1996 to 1998 (see references for details).

Key findings from TIMSS-94/95

In brief, the key achievement results pertaining to New Zealand students in 1994 were as follows.

Mathematics

- Both Years 4 and 5 students performed statistically significantly below the international means in mathematics. *Whole Numbers* was an area of relative weakness for these students.
- Both Years 8 and 9 students' mean achievement was just below the international means for these class levels but the difference was not statistically significant in either case. *Algebra* was identified as an area of relative weakness for these New Zealand students.
- On average, Maori and Pacific students' achievement was well below that of Pakeha/European and Asian students at each of the middle primary, upper primary, and lower secondary levels.
- In general there were small, non-statistically significant, differences between boys' and girls' mean achievement at the four class levels. However, when ethnic identity and gender were considered together, the low mean achievement for Maori boys at the middle primary level was of particular concern.

Science

- The achievement means of Years 4 and 5 students were equivalent to the international means for these two class levels. The area of *Physical Science* was identified as a relative weakness.
- The means for both Years 8 and 9 students were just above the international means. *Earth Science* was an area of relative weakness for these students.

[6] *Implemented at the beginning of the new school year in 1995.*

- On average, Maori and Pacific students' achievement was below that of their Pakeha/European and Asian counterparts at both the middle and upper primary, and lower secondary levels.
- At the lower secondary level, among participating countries, New Zealand observed one of the largest gender differences in mean achievement favouring boys. This difference was of statistical significance and largely attributed to the poor performance of Pakeha/European girls relative to the Pakeha/European boys.

Mathematics and Science Literacy[7]

- By international standards, New Zealand senior secondary school students performed statistically significantly above the international means in both the mathematics and science literacy domains.
- Large statistically significant gender differences favouring males in both literacy domains were observed.

The impact of TIMSS-94/95 in New Zealand

The TIMSS-94/95 findings for the upper primary and lower secondary were released in late 1996 and the middle primary school results in mid-1997. It was the release of the latter that highlighted areas of concern in mathematics and science education in New Zealand, particularly in the primary school sector. The TIMSS-94/95 results, together with reports that teachers, particularly in the primary sector, were experiencing difficulties in implementing the new curricula, were the impetus for the Minister of Education to establish the Mathematics and Science Taskforce in August 1997.

The Taskforce made a number of recommendations to which the New Zealand Ministry of Education has responded. These have included the establishment of a comprehensive publishing programme of new and revised resource material for both students and teachers, particularly those in primary schools. Other resources include a video and two CD Roms to support the new learning materials and the development of additional items for the online assessment resource banks. The publishing programme began in 1998 and is ongoing. In 1999, Te Kete Ipurangi (TKI), the Ministry of Education's online learning centre, was also launched as part of the 1998 Information Technology and Communications Strategy and is an example of another facility for supporting teachers in mathematics and science as well as other curriculum areas.

WHO CARRIED OUT TIMSS-98/99?

TIMSS-98/99 was administered under the auspices of the International Association for the Evaluation of Educational Achievement (IEA)[8]. Boston College's Lynch School of Education (in the United States) managed the international project activities. The other organisations that worked very closely with Boston College were:

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

The Comparative Education Research Unit, located within the Research Division of the Ministry of Education, was responsible for carrying out the TIMSS-98/99 activities in New Zealand.

[7] *The mathematics and science literacy study was administered in New Zealand during August, 1995.*

[8] *See Appendix A.1 for details on IEA.*

WHAT COUNTRIES OR EDUCATION SYSTEMS TOOK PART IN TIMSS-98/99?

Thirty-eight countries or education systems took part in TIMSS-98/99. Twenty-six of these countries also participated in TIMSS-94/95.

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

* Countries that also participated in TIMSS in 1994-1995.

THE CONCEPTUAL FRAMEWORK

The conceptual framework for both TIMSS-94/95 and TIMSS-98/99 was based on the concepts of the *intended*, *implemented*, and *attained* curricula.

The *intended* curriculum refers to the aims, content, and methods for each of mathematics and science education as defined by a country's educational authorities. The *intended* curriculum is described in documents such as curriculum guides, prescriptions, syllabuses, and policy statements. Textbooks, resources and examinations also reflect the essence of *intended* curriculum. As well as being set within a specific educational context, intended curriculum are also set within the context of a society. Societal factors — for example, the goals and expectations the society holds for education, the role of private education, the status accorded to teachers, the resources a society has, and the proportion of those resources it allocates to education — all have an influence on the national intentions.

Teachers interpret, translate, and implement the intentions of curricula according to their own experiences and beliefs. The educational milieu in which the *implemented* curriculum is placed embodies institutional arrangements made at the school and class levels but is also largely influenced by system level arrangements. The local community, while often reflecting society-at-large, provides the context for the setting of the *implemented* curriculum.

The *attained* curriculum consists of the concepts, processes, skills, and attitudes towards mathematics and science that students have acquired during their schooling. It can also be placed in the broader context of institutional arrangements as well as students' individual backgrounds (Robitaille et al, 1993; Robitaille & Maxwell, 1996).

CURRICULUM FRAMEWORKS FOR TIMSS-98/99

As noted at the beginning of the chapter, the frameworks developed for TIMSS-94/95 were used as a basis for developing the assessment test used in 1998/99 (see Figures 1a and 1b). Since about two-thirds of the TIMSS-94/95 test items used at the lower secondary level were released into the public domain, new items were developed as replacements. The frameworks were used to ensure that new items were similar in content, format, and difficulty to the items they were replacing[9].

[9] *In order to have new items for the final test, sets of potential replacement items were field tested in early 1998 in all participating countries. Twenty-seven New Zealand schools and approximately 1100 students were involved in the field trial.*

Figure 1a: The mathematics framework

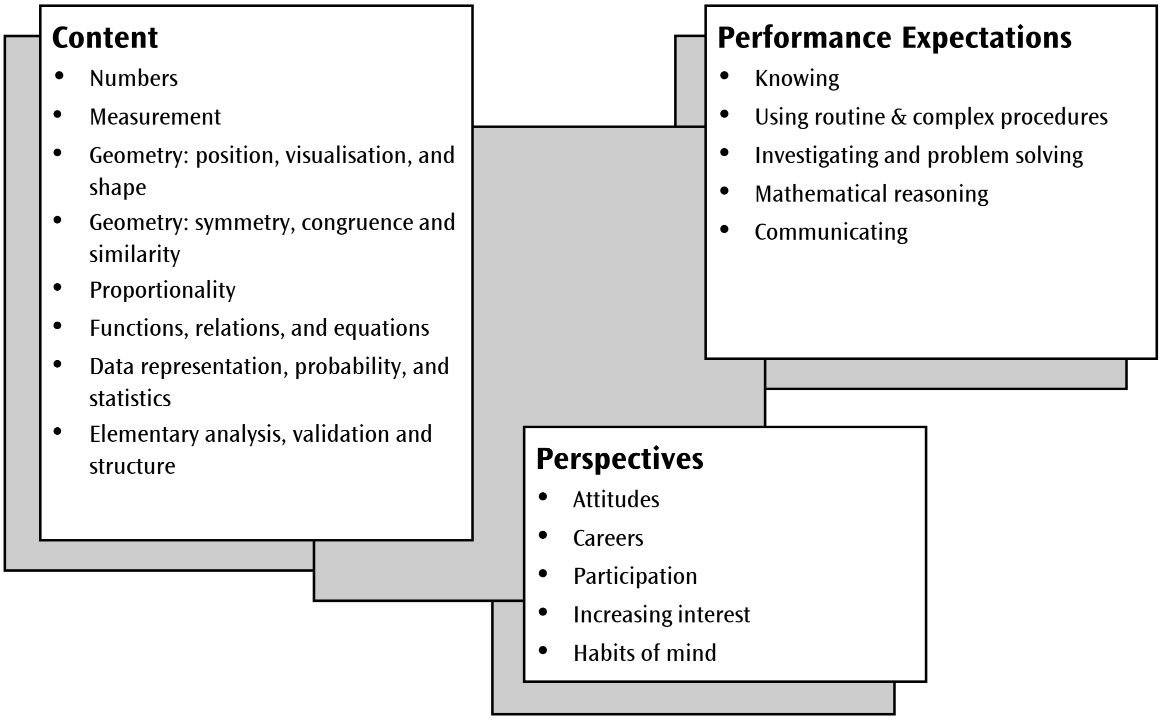
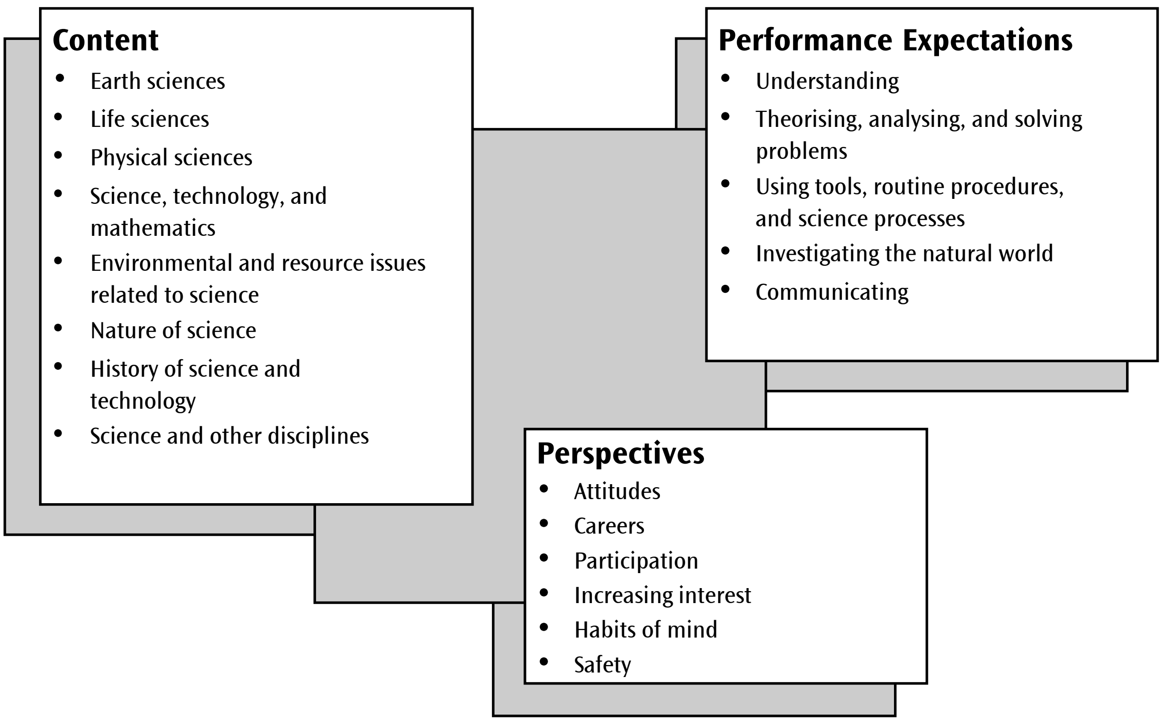


Figure 1b: The science framework



Source: Robitaille et al, 1993 (with minor adaptations).

WHAT STUDENT POPULATION WAS INVOLVED?

As noted at the beginning of this chapter, TIMSS-98/99 had a narrower focus than TIMSS-94/95 in that there was just one class level involved.

The international desired population for TIMSS-98/99 was:
All students enrolled in the *upper* of the two grades that contain the largest proportion of students of age 13 years (at the time of testing).

The New Zealand national desired population was:
All Year 9 (or form 3) students in October 1998.

Sample design

The sample design used in both TIMSS-94/95 and TIMSS-98/99 is generally referred to as a two-stage stratified cluster design. This involved:

1. selecting a sample of schools from all eligible schools; and

2. randomly selecting a mathematics (or science) class from each sampled school, regardless of the ability level of the class.

The main reason for selecting intact classes, rather than a random sample of individual students, was so that the beliefs and practices of teachers could provide contextual information around the achievement and attitudes of their students.

Countries were able to exclude schools if, for example, they were special education schools or were geographically isolated. Once schools and classes were selected, individual students could also be excluded if they had a physical disability which would prevent them undertaking the test in the specified time; they were notably intellectually or developmentally delayed; or if they had had less than one year's instruction in the language of the test. Since most countries achieved exclusion rates of five percent or less for TIMSS-94/95, the maximum exclusion rate was set at five percent for TIMSS-98/99. (For further details, see Martin, Gregory, and Stemler, 2000.)

Implications of the sample design

It is important to note that in both studies the achieved student samples *did not* come from using simple random sampling techniques. Therefore, it is not appropriate to use statistical formulae that are used when working with simple random samples. (See TN.1, TN.3, and TN.4 in Technical Notes for more details.)

TIMSS-98/99 was a population survey and the achieved sample of students was representative of the population from which it was drawn — in the case of New Zealand, Year 9 students. Similarly, the schools selected were representative of schools with Year 9 students. In contrast, the teachers who took part in the studies were not a representative sample of teachers; rather they were the mathematics and science teachers who taught a representative sample of students. (See Appendix A.2 for more details of New Zealand's sampling.)

MEASURES AND INSTRUMENTATION

Achievement test design

The design of the test and the approach taken to report estimates of achievement are very closely linked. As was the case in TIMSS-94/95, TIMSS-98/99 used scaling techniques based on Item Response Theory to generate the population estimates of achievement[10]. Such scaling techniques means that not all students need to be tested on the same set of items, enabling broad curricula coverage by the tests. These techniques mean the assessment is practical to administer, and enough accurate technical information can be generated to calculate achievement scores. A *multiple matrix sampling design* was used to construct the actual achievement test.

What does it mean to use a multiple matrix sampling design?

Essentially, a *multiple matrix sampling design* uses subsets of test items from a large pool of items that are then randomly administered to sub-samples of students. In practice this meant that the items used were assigned to clusters which were in turn assigned to at least one of eight booklets. A cluster could appear in more than one booklet. There was one cluster that was common to all booklets.

Each of the eight test booklets contained both mathematics and science items. Students were randomly assigned to one of these eight test booklets.

What was the format of the test items and how long was the actual test?

About 75 percent of the test items used in the study were in multiple-choice format and accounted for about two-thirds of the testing time. The remaining 25 percent were in free-response format (both short answer and extended response) and accounted for about one-third of the testing time.

The test was designed to take 90 minutes to complete, and was divided into two testing sessions of approximately equal length.

Student background questionnaire

Students completed a background questionnaire which sought demographic information, their attitudes to school, and to mathematics and science, and information about the types of after-school activities in which they were involved. The questionnaire took about 20 minutes to complete and was administered after the achievement testing session.

School and teacher questionnaires

The 'school' questionnaire sought information such as the size and location of the school, staffing levels, school leadership, and management practices, physical resources, and curriculum issues.

The mathematics and science teachers of the students who took part in the achievement testing completed questionnaires that asked for demographic details, for information about their academic and professional backgrounds, classroom practices, and about their attitudes towards teaching mathematics and/or science.

Each questionnaire took about 30–40 minutes to complete.

---

[10] *The underlying premise of Item Response Theory (IRT) is that there is a single underlying latent (or unobservable) trait on which students rely, to some extent, in order to give correct responses to a set of test items. Examples of such traits are verbal proficiency or mathematical facility. IRT assumes that it is "possible to describe mathematically the relationship between a person's trait level and performance on an item" (Stocking, 1999, pp55-63).*

Test administration in New Zealand

TIMSS-98/99 was administered in New Zealand at the beginning of the fourth school term in October 1998. Teachers in the participating schools administered the test.

REPORTING OF STUDENT ACHIEVEMENT RESULTS

In this report, achievement results are presented in the form of *Item Response Theory (IRT) scale scores*. The particular scale model used took into account such factors as item difficulty, item discrimination and (where applicable) the possibility of guessing, and provides better population estimates of underlying traits than is possible using a method such as percent correct. The scaling method also makes it possible to estimate distributions of proficiency in a population, as well as determine relationships between the scale scores and students' background characteristics, beliefs and attitudes. See TN.2 in Technical Notes for a summary outline. For a more comprehensive explanation of the scaling process the interested reader is referred to Martin, Gregory, and Stemler (2000).

It is important to note that the scale scores in this report which refer to TIMSS-94/95 are not the same scale scores as were reported in the earlier international reports for that study as the TIMSS-94/95 data have been rescaled. Because the original scale scores for TIMSS-94/95 were determined on the basis of results for two class levels, the achievement data had to be re-scaled using data from just the upper of the two class levels only, in order to make them directly comparable with the narrower population for TIMSS-98/99. In addition, a different mathematical model was used to scale the data this time around.

To aid interpretation of scale scores, a useful yardstick is that the results of TIMSS-94/95 and TIMSS-98/99 were placed on a reporting scale that had a mean of 500 and a standard deviation of 100.

Two other approaches were also taken to describe Year 9 students' achievements. These are as follows.

1. Reporting on how students have responded to individual mathematics or science items — for example, the percentage of students who answered each individual item correctly. This approach was used in TIMSS-94/95 as well; some examples are presented in Chapter 2.

2. Reporting a mean score for sets of items, commonly referred to as the mean percent correct score. This approach was used to report achievement results in the TIMSS-94/95 international reports and in New Zealand's national reports. This was the approach taken to report the preliminary national results, and was also used to examine trend results for each sub-topic area in mathematics and science.

Please note that this document reports basic descriptive analyses. Most of the analyses reported are bivariate (ie, they describe associations between two variables only) and do not, in themselves, demonstrate causal links. Rather, the report generally draws attention to similarities and differences between groups, and associations between achievement and other observed variables. The authors have also taken a relatively conservative approach in reporting statistical differences, which is also in line with the international reporting (see TN.4 and TN.5 in Technical Notes).

This chapter has provided a very brief outline of the background to the study, including the framework used in the study, as well as the curriculum frameworks used for the test item development. Perhaps one of the most complex aspects of the study was the development of the reporting metric for the achievement results. The methodology behind this work has only been touched on very briefly in this introduction. Chapter 2 examines New Zealand Year 9 students' achievement in mathematics and science in an international context.