Describing
Contexts, Curricula
and Achievement

Steven J May
Glenn S Chamberlain
Megan C Chamberlain
CHAPTER 3

CONTEXTS, CURRICULA, & ACHIEVEMENT

This chapter aims to provide a brief description of the New Zealand context in which TIMSS was conducted. In particular, this chapter will look at the content of the New Zealand mathematics and science curricula at the 9-year-old level, how this was intended to be taught, teachers and their training, resources for teaching, and assessment practices. Issues relating to changes in the curricula during TIMSS are also addressed. Steps in the development of TIMSS instruments to measure achievement and contexts are outlined.

The learning of mathematics and science is of fundamental importance to understanding the world we live in and functioning as an informed citizen and contributor to society. Politicians, educators, employers, and other social commentators have been unanimous in emphasising that New Zealand’s future economic development and competitiveness is dependent on our ability to comprehend and utilise mathematical and scientific ideas.

Increasingly rapid changes in technology have resulted in changes in the workplace. There will be a need for workers to be able to undertake regular retraining. Therefore, it is crucial that our students — tomorrow’s parents, workforce, and citizens — are adequately skilled in mathematics and science to cope with the technological advances.

MATHEMATICS AND SCIENCE CURRICULA

The Ministry of Education is responsible for determining both mathematics and science curricula intended to be taught. Individual schools, however, are free to include curriculum material that forms part of the needs and resources of the local community.

The mathematics and science syllabuses officially in use for standards 2 and 3 in the years prior to TIMSS were Mathematics: Junior Classes to Standard Four (Department of Education, 1985b) and Science Syllabus and Guide Primary to Standard 4 (Department of Education, 1979).

Mathematics Syllabus for Junior Classes to Standard Four

This curriculum statement provided the framework for planning and making decisions about a school’s mathematics programme, beginning with new entrants at age five. It states that mathematics encourages children to develop their curiosity about numbers, measurement, and
The aims of *Mathematics: Junior Classes to Standard Four* are set out below, namely for students to:

- develop a positive attitude towards, and a continuing interest in, mathematics;
- appreciate that mathematics is important in the world around them;
- develop the mathematical knowledge, skills and understanding required both for everyday living and as a foundation for further learning;
- discover mathematical patterns and relationships and form appropriate generalisations;
- develop an understanding of some of the principles underlying mathematics and the ability to apply them. (p6.)

The content of the syllabus was organised around nine topic areas:

- sets, relationships, and patterns;
- number;
- numeration;
- addition and subtraction;
- multiplication and division;
- money;
- geometry;
- measurement;
- graphs and statistics. (p18.)

Work on fractional numbers and decimals was included in the topics of number, numeration, and operations with number.

In addition to the topic areas, other key skills thought to be important for development at this level included:

- enquiry and inventiveness to search for solutions to answers;
- identifying applications of mathematics in the world around them;
- awareness of technological aids in mathematics;
- deriving satisfaction from mathematical activities. (p7.)

**Science Syllabus for Junior Classes to Standard Four**

This curriculum statement provided the framework for planning and making decisions about a school’s science programme. The *intended* primary curriculum included the introduction of children into the intellectual processes by which they acquired factual knowledge, concepts, skills, and attitudes. The study of science was said to have two main aspects — a method of enquiry and a body of knowledge.

The major aims of *Science Syllabus and Guide: Primary to Standard 4* are listed below:

- developing basic concepts;
- acquiring significant knowledge;
• developing an ability to communicate;
• developing process skills;
• developing interests and attitudes;
• developing an enquiring mind and the skills for exploring and interpreting the environment.

The content of the syllabus was organised around four broad topic areas:
• Living Things (living things; plants; animals);
• Matter (earth science; water; air; fibres; timber; glass; metals; plastics);
• Energy (mechanical energy and forces; sound energy; light energy; heat energy; electrical energy; magnetic force; chemical energy);
• Time and Space (time; space). (pp59–83.)

There are two other aspects on which the science curriculum placed much emphasis. Firstly, the development and application of process skills, such as observing, measuring, classifying, predicting, and experimenting. Secondly, the development of communication skills through oral and written presentations, reading, diagrams, tables of data, and graphs.

RECENT CHANGES TO THE MATHEMATICS AND SCIENCE CURRICULA

Mathematics

A redesign of the mathematics curriculum was initiated in 1991, as part of a broad review aimed at improving achievement in the essential subject areas. The new curriculum statement — Mathematics in the New Zealand Curriculum (Ministry of Education, 1992) — was introduced in 1993 at the primary level, although most schools had a draft version a year or so earlier. The new curriculum replaced the primary syllabus Mathematics: Junior Classes to Standard Four.

The new curriculum was written by a small project team under the guidance of a mathematics advisory committee. The views of schools and interested groups about the initial draft version were taken into account in producing the completed curriculum statement.

Changes in the content area of the new curriculum have been relatively minor. Strong encouragement to make use of new technology, in the form of calculators and computers, has attempted to address the advances occurring in this area.

One of the more significant differences occurred in the approaches recommended for teaching mathematics. The curriculum statement stresses the need for mathematics to be taught and learned within the context of problems which are meaningful to students and which lead to understanding of the way mathematics is applied in the world beyond. Other new influences include a greater focus on problem-solving strategies and the development of mathematical process skills. Finally, the new curriculum asks for an even greater effort in encouraging the participation and confidence of both girls and Maori children, and suggests methods to bring this about.
Mathematics in the New Zealand Curriculum re-emphasises the student-centred approach to teaching, and beliefs about teaching and learning, which have been a feature of New Zealand official education documents and teacher education for many years. Many of the facets which typify the stimulation of learning approach, as outlined in Chapter 2, are included. The curriculum aims to:

- help students to develop a belief in the value of mathematics and its usefulness to them, to nurture confidence in their own mathematical ability, to foster a sense of personal achievement, and to encourage a continuing and creative interest in mathematics;
- develop in students the skills, concepts, understandings, and attitudes which will enable them to cope confidently with the mathematics of everyday life;
- help students to develop a variety of approaches to solving problems involving mathematics, and to develop the ability to think and reason logically;
- help students to achieve the mathematical and statistical literacy needed in a society which is technologically oriented and information rich;
- provide students with the mathematical tools, skills, understandings, and attitudes they will require in the world of work;
- provide a foundation for those students who may continue studies in mathematics or other learning areas where mathematical concepts are central;
- help to foster and develop mathematical talent. (p8.)

Science

The science curriculum was also redesigned as part of the broad curriculum initiative in 1991. The new curriculum statement, *Science in the New Zealand Curriculum* (Ministry of Education, 1993), replaced the primary school syllabus *Science Syllabus and Guide Primary to Standard Four*. The schedule called for the implementation of the science curriculum in 1995.

The redevelopment of the science curriculum was managed by a small project team while a science advisory group provided guidance at key stages of the project. It builds on New Zealand research and curriculum development in science education.

In general, there were only minor modifications to the topic areas. One of the most significant changes was the inclusion of technology, in recognition of the need for students to be able to understand and cope with an increasingly technological environment.

The curriculum statement is said to make it possible to improve the learning process “by specifying clear learning goals expressed as achievement objectives”. Furthermore, it emphasises as important “for initial learning in science to be set within contexts which are meaningful to students and which lead to understanding of the inter-relationship of science, society, and technology” (O’Rourke in foreword to Ministry of Education, 1993, p5).

The aims of the science curriculum are to advance learning in science by:

- helping students to develop knowledge and a coherent understanding of the living, physical, material, and technological components of their environment;
- encouraging students to develop skills for investigating the living, physical, material, and technological components of their environment in scientific ways;
• providing opportunities for students to develop the attitudes on which scientific investigation depends;
• promoting science as an activity that is carried out by all people as part of their everyday life;
• portraying science as both a process and a set of ideas which have been constructed by people to explain everyday and unfamiliar phenomena;
• encouraging students to consider the ways in which people have used scientific knowledge and methods to meet particular needs;
• developing students’ understanding of the evolving nature of science and technology;
• assisting students to use scientific knowledge and skills to make decisions about the usefulness and worth of ideas;
• helping students to explore issues and to make responsible and considered decisions about the use of science and technology in the environment;
• developing students’ understanding of the different ways people influence, and are influenced by, science and technology;
• nurturing scientific talent to ensure a future scientific community;
• developing students’ interest in and understanding of the knowledge and processes of science which form the basis of many of their future careers. (p9.)

The new curriculum stresses that science is for everyone, regardless of their background. It notes the continued importance of encouraging the participation and confidence of both girls and Maori children.

**Curricular Changes and TIMSS**

Achievement data for TIMSS at the 9-year-old level were collected in 1994. The new mathematics and science curricula were officially introduced in 1993 and 1995, respectively. While some schools would have begun introducing the new curricula prior to these dates, the students’ achievement must be considered as being largely a product of teaching based on the old syllabuses. The new curricula would have had negligible influence on student achievement for TIMSS.

**International Trends in Mathematics and Science Curricula**

“Among the TIMSS countries, there was a widespread interest in reform of mathematics education: many reforms and managed changes were taking place in mathematics content, pedagogy, and technology use.” (Schmidt et al, 1997a, p33.)

The nature of the reforms varied across countries but common themes occurred across groups of countries. In Western countries, particularly, more emphasis was being placed on solving problems in ‘real-life’ contexts, communication in mathematics, and the use of calculators and computers as learning tools.

In science curricula, too, common reform themes emerged and these were similar in nature to the mathematics reform initiatives (Schmidt et al, 1997b). In the interests of deeper student engagement
and greater utility of the science taught, greater emphasis was being placed on teaching science in the context of everyday applications. Use of calculators and computers was being encouraged in many countries. Ecology, pollution, and the environment are topics that had been added to science curricula in several countries.

It should be noted that these reforms had been made to intended curricula and the extent to which changes had occurred at classroom level varied greatly.

Mathematics topics present in the intended curricula of at least 70 percent of the countries at the standard 3 level were:

“generally focused on number, measurement, and geometry topics. This included whole number ‘meanings’, whole number ‘operations’, ‘fractions’, measurement ‘units’, ‘perimeter, area, and volume’, and parts of two-dimensional geometry.” (Schmidt et al, 1997a, p18.)

The science topics most commonly included in the intended curricula (at least 70 percent of countries) for class levels equivalent to standard 3 are ‘bodies of water’; ‘human biology and health’; ‘weather and climate’; ‘Earth in the solar system’; ‘plants, fungi types’; ‘animal types’; ‘organs, tissues’; ‘interdependence of life’; ‘physical properties of matter’; ‘energy types, sources, conversions’; ‘land, water, and sea resource conservation’; and environmental issues such as ‘pollution’.

TEACHERS AND THEIR TRAINING

The teacher plays a major role in the development of children’s mathematics and science achievement. New Zealand teachers at the 9-year-old level are considered to be ‘generalists’ in that they must teach across a range of subjects, rather than specialising in one or two. In 1990, data collected from the IEA study on Reading Literacy showed that over 75 percent of standard 3 teachers were female, while almost all of the teachers had come from a background where English was the first language. The majority of teachers were also found to have completed 12 or 13 years of schooling, which equated to completing form 6 or 7, before undertaking teacher training (Chamberlain, 1993).

Training of primary teachers in New Zealand is undertaken in colleges of education. Currently, the majority of trainees undertake a three-year course leading to a Diploma of Teaching. This course can be reduced to two years for university graduates. New teachers then complete two years of satisfactory teaching to gain teacher registration.

Primary trainees are now encouraged to undertake university courses in conjunction with their college training — leading to a four-year Bachelor of Education degree. First started in the mid-1970s, this course is now available at all colleges of education.

The current three-year course requirements for trainee teachers were introduced in stages between 1966 and 1968. Consequently, teachers who completed their training prior to this period undertook two-year courses.
The number of core hours that all students currently receive on mathematics in college of education primary teaching programme varies between colleges of education, ranging from 50 hours to 100 hours. In comparison, minimum requirements for science are generally less than those for mathematics, although there are variations between colleges. Minimum core hours for science currently range from 36 hours to 96 hours. These times include instruction in both pedagogy and subject content. Trainee teachers may elect to spend more than the minimum times studying mathematics and science.

Teaching experience is an important factor in improving children’s achievement in mathematics and science. Chamberlain (1993) reported that the mean number of years of teaching experience for teachers of 9-year-olds in 1990 was 14.

CURRENT RECOMMENDED APPROACHES TO TEACHING

Mathematics

The new curriculum statement advocates a balanced mathematical programme which includes concept learning, developing and maintaining skills, and learning to tackle applications. According to the authors of the curriculum statement, students learn mathematical thinking most effectively through applying concepts and skills in interesting and realistic contexts which are personally meaningful to them. Thus, they believe mathematics is best taught by helping students to solve problems drawn from their own experience. Problem solving, it is argued, requires students to collect data, carry out experimentation, be flexible and creative, and evaluate. Such problems encourage thinking rather than just recall. Learning to communicate about and through mathematics is part of learning to become a mathematical problem solver and learning to think mathematically.

Garden (1997) has noted that classroom teaching in primary schools is inductive, that is, it is based on teaching of a range of specific examples upon which to build more general concepts.

Science

The science curriculum states that it is important for the school science programme to present a balance between knowledge and scientific skills.

Children’s learning of science is said in the curriculum statement for science to be most effective when they are taught scientific concepts and skills from real-life examples. This, it is said, provides the learner with a meaningful context within which to associate the new information. Children are then in a position where they may apply the new concepts and skills to more challenging situations.

Finally, learning is greatly enhanced when the concepts, ideas, and attitudes that children already possess are used as a starting point for learning (Ministry of Education, 1993).
TEACHING RESOURCES

Mathematics

*Beginning School Mathematics* (BSM) is a commonly used resource for junior classes in New Zealand primary schools. It was introduced in 1986 as a complement to the then new mathematics syllabus, *Mathematics: Junior Classes to Standard Four* (1985), and was designed to be used during the first three years of school. BSM is an activity-based resource comprising 12 cycles, with each cycle building on the content of the one before. BSM is very much a set of hands-on activities designed to enhance the children’s understanding of mathematical concepts and ideas (Visser, 1993).

Textbooks provide another resource for teachers. The new curriculum statement cautions against teachers following the content of textbooks too closely, since many contain material not applicable to the mathematics curriculum in New Zealand. In primary schools there is a standard textbook but this is commonly supplemented with topic booklets or other teacher-made material (Garden, 1997).

Science

An extensive series of small topic booklets was developed and published by the then Department of Education during 1980–1982, and these are still in use. The aim was to provide teachers with a resource to achieve the aims and objectives of the New Zealand syllabus for the teaching of science in the primary schools. There are three broad age ranges for which each series is designed. Each booklet focuses on one of four topic areas, as well as the development of understanding and process skills required for scientific thinking. The use of teacher-produced material is also a common feature at the primary school level.

Teachers generally use textbooks to support their classroom science programme, rather than as a basis for the programme (Garden, 1997).

CLASS ORGANISATION

A common characteristic of primary schools in New Zealand is the use of composite classes, that is, a class in which two or more grade levels are taught by the same teacher. In the Reading Literacy Study, about three-quarters of the teachers of 9-year-olds, teaching 61 percent of students, reported that they taught composite classes (Chamberlain, 1993).

From the same data, the mean class size at the 9-year-old level was just under 30.

ASSESSMENT PRACTICES IN MATHEMATICS AND SCIENCE

The new curriculum statements stress that evaluation is necessary to assess children’s readiness for new learning, to give teachers feedback on the success of their methods and approaches, and to assist
planning for new learning. According to the curriculum statements, evaluation should include the use of diagnostic testing to enable teachers to be aware of any potential problems; and classroom assessment should include a range of methods including written, oral, demonstration, and group formats. Finally, it is seen as important that evaluation provide students and their parents with a summary of a student’s progress.

A major change in emphasis between the new curricula and former syllabuses is the inclusion of achievement goals at each class level. This is intended to allow primary schools to relate students’ progress to nationally defined achievement goals.

*Beginning School Mathematics* provides teachers of junior classes with opportunities to monitor student progress. The assessment checkpoints are closely matched to the teaching objectives and involve teachers observing and listening to each child and keeping records on these observations (Kirk & Gilmore, 1995).

A standardised test with national age and class norms, the *Progressive Achievement Test of Mathematics (Revised)* (1993), is produced by the New Zealand Council for Educational Research (NZCER). First published in 1974, the PAT Mathematics is designed for use with classes from standard 2 to form 4. It is normally used at the start of the school year to provide teachers with information on students’ abilities, to assist them in making classifications for classroom instruction. In addition to mathematics, there are also the *PAT Reading Comprehension (Revised)* (NZCER, 1991a) and *Reading Vocabulary (Revised)* (NZCER, 1991b) which are used widely by teachers (Croft and Reid, 1991). These tests, which use a multiple-choice format, along with other classroom assessment, mean that most 9-year-olds are familiar with a variety of question formats.

The *Assessment Resource Banks* are another recent NZCER development. The initial focus is on mathematics and science. Each item covers a part of the national mathematics or science curricula. While there is a wide range of item formats including multiple-choice, short-answer, oral, open-ended, practical tasks, questionnaire and interview schedules, only reliable and valid items are included. Another important feature is that it provides user-friendly access for teachers (Reid et al, 1994).

**TIMSS TESTS AND QUESTIONNAIRES**

To obtain high quality data for analysis, TIMSS has put considerable resources into the design and development of the tests and questionnaires. Procedures used in the administration and data processing phases of this study have also contributed to the quality of the data. This section and subsequent sections describe both the development and the implementation processes for the study.

An international perspective of the study design and development of instruments can be found in the second TIMSS monograph *Research Questions & Study Design* (Robitaille & Garden, 1996). More detail about the development and implementation of the study is reported in the international *TIMSS Technical Report Volume 1* (Martin & Kelly, 1996). Detail about New Zealand implementation will be in the New Zealand technical report for TIMSS (May & Udy, in press).
What the Tests and Questionnaires are Designed to Measure

The tests and questionnaires have been designed to provide the data necessary to attempt to answer four general TIMSS research questions. As noted in Chapter 1, one of these questions is “What mathematics and science concepts, processes, and attitudes have students learned; and what factors are linked to students’ opportunity to learn?” To answer this question, measures of student achievement that were both valid and reliable were needed to assess the extent of learning of concepts, skills, and processes in mathematics and science. Student opinions on issues related to learning mathematics and science were needed to help determine attitudes and beliefs. A variety of background and contextual questions for students were needed to help find the important factors linked to students’ opportunity to learn and achievement.

Another question “How are the intended, the implemented, and the attained curricula related with respect to the contexts of education, the arrangements for teaching and learning, and the outcomes of the educational process?” is also asked. To help answer this question, principals and teachers of sampled students answered questions designed to set the context of the students’ mathematics and science learning. When this is analysed in conjunction with the data obtained from the students and system-level questionnaires, relationships between the intended, the implemented, and the attained curricula can be investigated. Interpretation of these analyses will be aided by the intensive analyses of curriculum documents and commonly used textbooks in mathematics and science.

Achievement Instruments

A decision was made at the outset of the study to use a variety of item types to measure student achievement. The item types utilised in this study include multiple-choice items, free-response items, and performance tasks. Every student answered multiple-choice and free-response items as part of a written testing session. In this testing session, each student was administered one of the eight achievement test booklets assembled for population 1. These booklets contained both mathematics and science items.

As with many previous IEA studies, TIMSS has used a large number of multiple-choice items. Use of these items enables a large number of students to be tested over a wide range of topics. Tests constructed using multiple-choice items generally have high reliability and high validity for many important measurement objectives.

To cover achievement outcomes that are either impossible to measure or very difficult to measure using multiple-choice items, three other item types were employed. The free-response items were either short answer, requiring a simple calculation or one- or two-word response, or were extended response items. The extended response items tended to be more complex and could, for example, involve students doing a series of calculations for a mathematics item, or writing a paragraph or two to explain a process for a science item. Performance Assessment tasks were designed to test skills and outcomes that are not easily assessed using pencil and paper tests.

A sub-sample of the standard 3 students was selected to take part in Performance Assessment on a separate day from that on which they answered the pencil and paper test. The hands-on tasks were
administered in an arrangement where students attempted tasks at three of the nine stations in the circuit. Some of the tasks involved mathematics activities, some science activities, and some a combination of both mathematics and science activities. Preliminary results from Performance Assessment were reported in Caygill (1995).

Student Questionnaire

All students were also required to complete a student questionnaire. From this, background data on students were obtained. These data included their age, gender, the ethnic group(s) they identified with, and whether they were born in New Zealand. Data about their family environments were collected; for example whether English was spoken at home, how many people lived at home and who they were, how many books there were in the home, whether they had various non-essential possessions in the home1, and the importance that their mothers placed on academic achievement. Data on the types and frequency of out-of-school activities were also collected. Students’ attitudes to mathematics and science were characterised from their responses to questions asking whether they agreed or disagreed with various statements. Students were also asked about learning activities that went on in their mathematics and science classes.

Teacher Questionnaire

Teachers of the students selected in the sample completed a teacher questionnaire. At the population level, there was one questionnaire which covered both mathematics and science teaching. Data were collected on the teachers’ academic and professional backgrounds; their instructional practices; and their attitudes towards, and beliefs about, teaching mathematics. Data were also collected on the class or group that the sampled students belonged to; for example, teachers’ assessments of the ability of their class. For both mathematics and science, teachers were asked about what factors limit their teaching to their particular group and what topic areas had been covered by the class. In addition, for mathematics, teachers were asked about textbook usage, calculator use, lesson preparation, content and structure of the last mathematics lesson, teaching strategies, and homework. These and other data provide important information on the implemented curriculum.

School Questionnaire

Principals of the schools involved in the study completed a school questionnaire. This questionnaire sought background data on the school, staffing, and the student population as well as the schools’ policies. Information on how mathematics and science curricula were implemented, availability of resources, provision of in-service training for staff, and staff retention rates were of particular interest.

1 Intended to contribute to a surrogate measure for socio-economic status.

47
THE DEVELOPMENT PROCESS

Item Selection

Items for use in the TIMSS achievement test came from a variety of sources. Items from this pool were fitted to mathematics and science test blueprints (McKnight et al, 1993) based on curriculum frameworks developed for this study (Robitaille et al, 1993). Items were drawn from the various content areas and with differing performance expectations to sample the universe of mathematics and science curricula. At later stages of development, additional items were added to ensure that relevant content areas and performance expectations were adequately covered. Some of these items came from national centres and some were produced to specification for TIMSS by other agencies experienced in large-scale assessment.

Each item was to be appropriate for almost all countries and cultures, although the design also had to include items that allowed the examination of differences between national curricula. A balance had to be struck between these two objectives.

Consequently, the items went through a lengthy process of review, with many items being changed or discarded at various stages of the process.

Bias

One issue of concern was that of possible bias in the content or context of an item. The context of a question should not unfairly disadvantage students from any particular country or group of countries, or comprise a significant proportion of the student population of a country (eg gender, ethnic, or religious groups). For example, items referring to a particular animal species may advantage students from countries in which the species is common and disadvantage students from countries where it is not found. Some items were rejected for this sort of reason. Other items were reworded or adapted for TIMSS. For example, a general question could ask the student to give a specific example familiar to them in their response.

Another screening device involved the use of panels of subject-matter experts to further refine the pool of items. These experts also made suggestions on how best to augment the pool when there were shortages of items in certain areas.

Item Trialling

The actual performance of items under TIMSS testing conditions also assisted with the process of item selection. Items were trialled internationally by means of a carefully controlled item pilot in April/May of 1993 and in a field trial2 of TIMSS instruments and procedures in early 1994. The

---

2 The field trial was a trial of all the procedures (sampling, test administration) to be used in the main survey.
analysis of item statistics obtained from the item pilot and field trial helped determine which items were appropriate for the majority of participating countries.

Items were trialled in 43 countries in April/May 1993. Each participating country was expected to trial each item on at least 100 students, to enable stable item statistics to be calculated. In New Zealand, as in most countries, this target was exceeded. In addition, the items were reviewed at the national centres in terms of their appropriateness to the curriculum and a judgement made on the quality of the item. National centres also made suggestions on how to improve items, particularly with regard to language used.

The results of the item pilot and subsequent review by subject-matter experts indicated that the item pool needed augmentation. Items added to the pool were able to be trialled during the field trial in early 1994, which also enabled countries to carry out a trial of sampling and test administration procedures to be used in the main survey. From this field trial, and with a more detailed review of the items by national centres and experts, the final item selection for the main survey was made. Final achievement tests were approved by national research coordinators from all participating national centres.

A more detailed account of development of the TIMSS achievement tests is given in Garden and Orpwood (1996).

**Questionnaire Development**

Another important facet of the preparation of instruments was the development of appropriate context questionnaires to enable meaningful data to be collected from school principals, teachers, and students in all countries. These questionnaires were developed with the conceptual framework of TIMSS in mind. Questions were developed to assist in answering the four general research questions, as well as to answer the specific questions relating to these general questions.

The questions for the context questionnaires were reviewed at national centres, and again these questionnaires were trialled to assess the appropriateness of the component questions internationally. Student questionnaires were trialled along with the item pilot and field trial. A separate trial of teacher and school questionnaires was held in October/November 1993. Data gathered by means of these questionnaires is discussed in later chapters of this report.

**NEW ZEALAND INVOLVEMENT IN THE DEVELOPMENT PROCESS**

New Zealand was involved with most phases of the international development process. Staff from the IEA Unit carried out most of the operational work involved. Additional assistance and input came from members of the National Advisory Committee (NAC) (see Appendix 2). This committee consisted of teachers from primary, secondary, and tertiary sectors, some of whom represented New Zealand Education Institute, Post-Primary Teachers’ Association, and the Independent Schools
Association; Ministry of Education and other curriculum and policy experts; and a School Trustees Association representative. At meetings of the NAC, members advised on a wide range of TIMSS instruments and procedures for implementation in New Zealand.

Data collection began in July 1991 with the first participation questionnaire, which sought general information about the New Zealand education system such as the structure of the school system, student roll numbers by age and class level, and length of the school year. This information was critical for matters of population definition, sampling, and survey timing and administration. The second participation questionnaire, completed in September 1991, looked at further data in relation to the location of decision-making on matters of curriculum and instruction, what qualifications were needed to become a teacher, and course structures for mathematics and science.

During 1992 and 1993, staff from the IEA Unit undertook two of the Curriculum Analysis strands referred to in Chapter 1 — Document Analysis and the General Topic Trace Map. Document Analysis involved a detailed analysis of the curriculum guides and a selection of textbooks. As New Zealand was in a transition from the former mathematics and science curricula to new curriculum statements, both the former and new curriculum guides were analysed. Textbooks analysed were those considered to be in common usage from information obtained from textbook sales and from teachers. General Topic Trace Map required a sequence analysis of each country’s mathematics and science curricula, locating initial and final stages of instruction, and the levels at which study was concentrated. The international results from the Curriculum Analysis component of TIMSS are reported in Schmidt et al (1997a & 1997b)

National Item Trialling and Review

In New Zealand, 1437 primary school students and 2568 secondary school students took part in the item pilot during April 1993 (Caygill & Chamberlain, 1993). This number of students enabled trialling of 1095 achievement items distributed over 32 booklets. Each student answered the items in the one test booklet assigned to them randomly. This resulted in between 113 and 132 students answering every item, which meant that reliable statistics for each item could be calculated to provide data to assist with item selection for the main survey.

In addition, each test booklet included a set of questions to capture information about the student’s background. These were being trialled for the student questionnaire to be used in the main survey.

As part of the item pilot, each item was reviewed by a panel. The review looked at whether the content tested by the item was appropriate for standard 3 or form 3 students. New Zealand reviewers also gave recommendations over which items they believed should be included, modified, or deleted from the item pool.

As mentioned previously, the trial of the teacher and school context questionnaires took place in October 1993. Data were obtained from 45 teachers at the standard 2 and standard 3 level, and from

---

3 Twelve booklets at the primary level and 20 at the secondary level. This large number of booklets (and students) was used to minimise the testing load on any one student.
49 teachers at the form 2 and form 3 level. Responses were also sought and received from 24 principals from these two levels, and five principals at the form 6 and form 7 level. Although the number of individuals involved was small, the intent of this pilot was to examine question content and format, and for this the comments supplied by the participating teachers and principals were very helpful. As a result of this pilot, the staff of the IEA Unit were able to supply many recommendations as to the content and wording of the questions (Chamberlain, 1994a).

New Zealand did not take part in the field trial for populations 1 and 2 as this was scheduled in the same school year as the main survey in the Southern Hemisphere. However, the items from the field trial were reviewed by the national centre in Wellington. This review was more detailed than the review for the item pilot, with a rating given on five specific questions for each item. Reviewers were provided with detailed explanations as to how the questions were to be interpreted.

The specific questions were:

1. Have the target students in your country been taught the content tested by this item and, if so, how much emphasis was put on it?

2. To what extent is the item similar to typical items to which your students are exposed when this topic is taught and/or tested?

3. How difficult do you think this item will be for target students in your country?

4. How would you rate the overall appeal of this item?

5. Do you consider this item culturally biased and hence unfairly disadvantageous to many or all of your target students?

Source: Garden, 1996c.

The ratings and recommendations given to each item, particularly regarding wording, were submitted to the TIMSS Study Centre. The New Zealand review, along with the review and the field trial data from other countries, went into the process for the final item selection.

In April 1994, New Zealand also took part in the population 3 field trial which helped refine administration procedures to be used in the main survey (Chamberlain, 1994b).

As a result of having participated fully in the item pilot, context questionnaire trial, and population 3 field trial, the IEA Unit staff had knowledge of and experience in the content and administration of nearly all of the achievement items and context questions to be used for the main survey. Throughout the development period, the IEA Unit was in regular electronic mail and facsimile contact with the international centres for the study in Vancouver and Boston, and the unit members were able to make a considerable contribution to the international development of TIMSS.
THE MAIN SURVEY INSTRUMENTS

The Final Item Pool

As a result of the development process, the final pool of items contained items that were individually approved by the national research coordinators, had been proven under test conditions, and covered the range of content and performance expectations as specified by the international test blueprints.

The items in the population 1 test came from 11 mathematics content areas and seven science content areas. Each item also had a performance expectation associated with it. The performance expectation categories for mathematics and science are described in detail in Chapters 4 and 5, respectively. The number of testing minutes for each content area and performance category for mathematics is shown in Table 3.1 and for science is displayed in Table 3.2.

Table 3.1

Mathematics: composition of test by content
and performance expectation (minutes)

<table>
<thead>
<tr>
<th></th>
<th>Knowing</th>
<th>Routine</th>
<th>Complex</th>
<th>Solving</th>
<th>Justifying and Proving</th>
<th>Communicating</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Numbers:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· place value</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>· other content</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Decimal Fractions</td>
<td>3</td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Common Fractions</td>
<td>4</td>
<td></td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Proportionality</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
<td>6</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Estimation and Number Sense</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Measurement</td>
<td>5</td>
<td>6</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>3</td>
<td>8</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Geometry</td>
<td>9</td>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Patterns, Relations, and Functions</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>42</strong></td>
<td><strong>16</strong></td>
<td><strong>26</strong></td>
<td><strong>21</strong></td>
<td><strong>3</strong></td>
<td><strong>10</strong></td>
<td><strong>118</strong></td>
</tr>
</tbody>
</table>

*Source: Adapted from Adams & Gonzalez, 1996.*
Table 3.2

<table>
<thead>
<tr>
<th></th>
<th>Understanding</th>
<th>Theorising</th>
<th>Using</th>
<th>Investigating</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Earth Science:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· earth features</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>· other content</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td><strong>Life Science:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>· human biology</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>· other content</td>
<td>31</td>
<td>1</td>
<td>1</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td><strong>Physical Science:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td><strong>Environment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td><strong>Other Science Content</strong></td>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>84</td>
<td>24</td>
<td>6</td>
<td>3</td>
<td>117</td>
</tr>
</tbody>
</table>

Source: Adams & Gonzalez, 1996.

Overall, the multiple-choice items make up over three-quarters of the test items (65% of the testing time), with free-response questions making up the remainder of the items (35% of the testing time). For population 1, the multiple-choice items were estimated to take one minute to complete. Free-response items were estimated to take one minute for short answer items and three minutes for extended response items. Times allowed were generous because students from some countries tend to take longer to do tests than those from others and there was a desire to avoid the test being ‘speeded’.

Common items were included in the different test instruments for each population to enable comparison of the achievement and response patterns of different age groups on certain items. There were 32 items in common between population 1 and 2, of which 15 were mathematics items and 17 were science items.

**Layout of the Test Booklets**

In order to maximise coverage of both mathematics and science content and the performances expected of students in the limited time schools could spare for achievement testing, a multiple matrix sampling design was used. This meant that instead of each student doing all the achievement items for mathematics and for science, they would only do a sample of these items. This would enable the testing load for an individual student to be limited to an acceptable level. To accomplish this, the items were arranged into test booklets.
The test booklets were designed so that several criteria were met:

• total testing time for any student was to be no longer than 70 minutes for population 1;
• items were to be distributed amongst booklets according to a design which would maximise reliability of the test and sub-tests to be reported;
• the booklets were to be approximately equal in difficulty;
• the booklets were to have both mathematics and science items although not necessarily in equal proportions in every booklet;
• every booklet was to include both multiple-choice and free-response items.

Source: Adapted from Adams & Gonzalez, 1996.

To achieve this, the items were grouped into mutually exclusive clusters. Each item cluster was designed on the basis of the estimated time to complete rather than number of items. This resulted in 26 clusters of items, with one of the clusters estimated to take ten minutes to complete, while each of the remaining 25 clusters were estimated to take nine minutes to complete for population 1.

The clusters themselves could be categorised into six different types:

1. The core cluster containing five mathematics and five science items which were all multiple-choice. This cluster appears in the second position in each test booklet.

2. Seven focus clusters with mathematics and science items. These clusters contained multiple-choice items. Focus clusters were assigned to booklets using a balanced incomplete block design. This has the effect that each cluster appears in three booklets but in different positions.

3. Five mathematics breadth clusters with an estimated nine minutes of mathematics items. Both multiple-choice items and free-response items were used for these clusters. Each breadth cluster appears in one booklet only.

4. Five science breadth clusters with an estimated nine minutes of science items. Both multiple-choice items and free-response items were used for these clusters. Each breadth cluster appears in one booklet only.

5. Four mathematics free-response clusters which each appear in two booklets.

6. Four science free-response clusters which also each appear in two booklets.

Source: Adapted from Adams & Gonzalez, 1996.

These clusters were then assigned to booklets. Seven of the booklets had the following structure:

<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Focus</th>
<th>Core</th>
<th>Focus</th>
<th>Free-Response</th>
<th>Focus</th>
<th>Breadth</th>
<th>Free-Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position in booklet</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

54
The eighth booklet had the following structure:

<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Focus</th>
<th>Core</th>
<th>Breadth</th>
<th>Free-response</th>
<th>Breadth</th>
<th>Breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position in booklet</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

To try to minimise the effect of fatigue on students’ performance, each booklet was organised into two testing sessions. The first clusters in a booklet were administered in a 37-minute session and the remaining clusters in a 27-minute session, with a break of up to 20 minutes between sessions.

**Questionnaire Structure**

The final questions selected for the school, teacher, and student context questionnaires survey were generally arranged in blocks of related questions. For example, the student questionnaire began with demographic questions about the student, followed by questions on student use of time, then questions about the family environment. Further blocks of questions also included questions on student attitudes to mathematics and science, and questions on what activities occur in mathematics lessons.

This grouping of similar questions helped assist respondents to remain correctly focused on the intent of the questions, as well as providing cross-validation of some of the responses.

**Translation/Editing**

Approximately two months prior to administration, the final versions of the test booklets, questionnaires, and administration manuals were received from the TIMSS Study Centre. Each document required editing to ensure that each item, question, or instruction was appropriately phrased for the New Zealand context. In some cases this involved changing words or phrasing that were unfamiliar to New Zealand students. For example, ‘petrol’ was substituted for ‘gasoline’. Any changes to the achievement items had to be approved by the TIMSS Study Centre to ensure that no substantive changes had been made which would affect the comparability of New Zealand data with that from other countries.

**QUALITY CONTROL FOR ADMINISTRATION OF TESTS AND QUESTIONNAIRES**

To ensure that data collected in each country participating in TIMSS is comparable, administration procedures and documentation were developed by the TIMSS Study Centre in Boston. In each country, the national centre overseeing the study was responsible for ensuring adherence to these standard procedures. An important way of safeguarding the strict international procedures was through the liaison between a participating school’s nominated contact person (‘school coordinator’) and the national centre (in New Zealand’s case, the IEA Unit). School coordinators were responsible for supplying the IEA Unit with all data needed for class level sampling, as well as making time available for meeting with IEA Unit staff to go over the testing procedures. The school
The Test Administrator Manual documented all procedures to be used in all countries at the time of testing. In many cases this was the school coordinator or the classroom teacher, but could be any responsible person briefed in the administration of the test materials. An 0800 help line was set up to enable school coordinators to contact the IEA Unit to get a rapid response to any queries they might have. It was essential that the timing for the testing sessions was strictly adhered to. Testing was carried out under normal testing conditions similar to those that many students would be familiar with for Progressive Achievement Tests. The test administrator was not allowed to answer any questions about the content of the test items but could clarify instructions. The instructions for testing were to be read from the Test Administrator Manual verbatim to ensure that testing sessions were administered in the same way to all students, both within New Zealand and in all participating countries.

The student questionnaire was administered in a further session. Twenty minutes was allowed for this, but this was extended if students required more time to complete the questionnaire. The test administrator assisted students who required help with any of the questions.

**Use of Calculators**

Although the use of calculators is encouraged by many teachers in New Zealand, calculators were not permitted for population 1. This was for several reasons. Firstly, the administration had to be the same in all countries world-wide. Secondly, some countries do not permit the use of calculators at the population 1 level and in some countries students would not have access to calculators. Thirdly, students could be tested on their ability to carry out mathematical operations (without the use of a calculator) and their ability to estimate could also be measured.

In addition, the content of some of the estimation items also meant that measuring instruments such as rulers were not permitted.
Testing Dates in New Zealand

The administration of the pencil and paper test took place between 17 October and 28 October, 1994. School coordinators were advised not to test on the days either side of a weekend to minimise the effects of absenteeism. Performance Assessment took place with a sub-sample of the students between 17 October and 25 October on a different day from the pencil and paper test.

Quality Assurance

To further assure the quality of the data collection, the Quality Assurance Programme was developed to document TIMSS activities amongst the participating countries. Each country was required to nominate a person independent of the IEA national centre to be responsible for monitoring and reporting on administration of the study. This included site visits to schools when the testing was taking place, an interview with the school coordinator, and an interview with the national research coordinator. The quality assurance monitors then made a report on the national content of the study to the TIMSS Study Centre. At the time of the data collection for populations 1 and 2 in New Zealand, the Quality Assurance Programme was still being developed. It was in place for the Northern Hemisphere’s data collection in April–May 1995, and therefore was used during New Zealand’s population 3 data collection in August 1995.

To follow up on the actual administration, test administrators completed a form that reported on the timing of the test sessions and on any deviations from procedures or unusual circumstances that may have affected the testing sessions. These forms indicated that all went well with the testing sessions in New Zealand. It was noted that nearly all students were able to complete the tests within the time allowed.

QUALITY CONTROL FOR DATA RETURN AND PROCESSING

After the testing sessions were completed, school coordinators packaged the materials and returned them to the IEA Unit. The materials were checked and prepared for coding of the free-response items and for data entry. The quality control procedures used in these processes are described below.

Coding of Free-response Items

After the data were collected, and prior to data entry, the free-response items required coding. Each free-response item was coded using a two-digit code, with the first digit indicating correctness and the second digit indicating the method or approach to the question used by the student. The coding scheme will be outlined in a supplementary publication on free-response items.

The free-response codes were developed as part of the piloting and trialling procedures. International training sessions were used to help ensure that these questions were coded uniformly in
all countries. One of the New Zealand IEA Unit staff attended an international training session held in New Zealand, along with IEA staff from Australia, Singapore, and Korea.

The internationally trained coder then convened training sessions for the New Zealand coding staff. Training involved tasks such as marking of sample items, discussion of rationale for marking, and comparison with international benchmarks. During coding of the free-response items, the coding staff met frequently to discuss problem responses and clarify codes used for marking.

As part of the quality control process for the free-response questions, a random sample of 10 percent of student booklets was marked by two coders. Analysis of this double marking for 28 mathematics items showed the codes exactly matched for between 88 percent and 100 percent of cases, with most (25) of the items exactly matching in over 90 percent of cases. Similarly, 27 science items showed the codes exactly matched for between 83 percent and 100 percent of cases, with most (21) of the items exactly matching in over 90 percent of cases. Between 92 percent and 100 percent of cases matched on the initial digit, which indicated the correctness score for the item. All items were above the 70 percent minimum standard set internationally for within-country, inter-rater agreement of free-response coding.

An international coding reliability study for population 2 was designed to verify how well countries fared in achieving consistent marking for the free-response questions. Analyses of data collected from the study indicated that rater agreement was very high across countries (Mullis & Smith, 1996). Given that the same procedures for coding were used for population 1 and the high inter-rater agreement for population 1 in New Zealand, it would suggest that reliability was also high for the coding of population 1 free-response items across countries.

**Data Entry**

Data entry provided a further opportunity to check the data provided by respondents and entered by the data entry personnel. The data entry program was custom written for this study and provided checks on out-of-range values; some checks on logical inconsistencies between different responses; a check to ensure that column shifts did not occur; checks on identification variables; and a check on the linkages between school, teacher, and student data files. Most problems were resolved during data entry.

Also during data entry, a 10 percent sample of respondents for each data file had their data double entered. The error rate for the double entry generally ranged between 0.4 percent and 0.6 percent, which was well within the international standard of one percent set in the international administration manual for the study.

After the data had been entered, the files were submitted to the international data processing centre in Hamburg for further cleaning and processing. New Zealand data required very little in the way of further checking.
SUMMARY

The TIMSS study involved an extensive development process. Each participating country provided data at various stages of this process to help ensure that the instrumentation for the main survey would assist with answering the TIMSS research questions. Of particular importance was the development of achievement tests. Each item has been reviewed for technical aspects (eg for wording and for test performance measured by item statistics), and also, as to how well the item fits the countries’ curricula.

The test booklets used for the main survey contain the best selection of items that were each appropriate in most countries, proven under test conditions, and cover a range of content areas and performance expectations.

Various methods were used to make sure that the quality of the test administration and the quality of the data collected were high. Emphasis was placed on the importance of following standard procedures for the test administration, and data collected were subject to rigorous checking and quality control procedures.

Student, teacher, and school data for population 1, the population under study in this report, were collected in October 1994, with testing for the majority of schools being conducted in the last two weeks of that month.