CHAPTER 2

TIMSS design and methodology

This chapter will provide an overview of the design of the study. TIMSS was a large-scale comparative study involving 50 participants. TIMSS primarily measured learner achievement in mathematics and science, as well learner beliefs and attitudes towards these subjects. The study also investigated curricular intentions and school and classroom environments. Since TIMSS was an international exercise, there was a need for a common framework to ensure comparability of the results across the different countries, and a need to develop instruments that would be useful to each of the participating countries. To this end, TIMSS International Study Center (ISC) organised a number of meetings with the national research co-ordinators to seek consensus on the framework and items to be included in the various instruments. In addition, TIMSS produced manuals to assist in administrating the study in the different countries.

This chapter will provide a description of the study's design and framework. This description is drawn mainly from two TIMSS publications: *TIMSS Assessment Frameworks and Specifications 2003* (Mullis et al. 2003) and *TIMSS 2003 Technical Report* (Martin, Mullis & Chrostowski 2004). The chapter will also describe how the study was conducted in South Africa.

TIMSS conceptual framework

TIMSS uses the curriculum, broadly defined, as the organising principle in how educational opportunities are provided to learners. The curriculum model has three aspects: the *intended curriculum*, the *implemented curriculum* and the *attained curriculum*.

The *intended curriculum* refers to the mathematics and science knowledge that society intends learners to learn, that is, the curriculum at the national or system level. Information from this level provides a national, social and an educational context. The *implemented curriculum* refers to how the educational system should be organised to facilitate this learning; what is actually taught in classrooms, who teaches it, and how it is taught. Data collected here provides information on the school, teacher and classroom context. The *attained curriculum* refers to what it is that learners have learned and what they think about these subjects. Data collected here provides information about learner outcomes and characteristics.

Instruments

Using the above model, TIMSS used mathematics and science achievement tests to describe what learners have learnt. The learner, teacher, and principal questionnaires obtained information on the structure and content of the intended curriculum in mathematics and science; the preparation, experience and attitudes of teachers; the mathematics and science content actually taught; the instructional approaches used; the organisation and resources of schools and classrooms; and the experiences and attitudes of the learners in the schools.

Achievement instruments

TIMSS assessed in the areas of mathematics and science. TIMSS 2003 was framed by two organising dimensions: a content domain and a cognitive domain. The content domain defined the specific mathematics and science subject matter covered by the assessment, while the cognitive domain defined the set of behaviours expected of learners engaged in mathematics or science.

The content domains that framed the mathematics curriculum were: number, algebra, measurement, geometry and data. The cognitive domains for mathematics were: knowing facts and procedures, using concepts, solving routine problems, and reasoning. The content domains that framed the science curriculum were: life sciences, chemistry, physics, earth science and environmental science. The cognitive domains were: factual knowledge, conceptual knowledge, and reasoning and analysis. Table 2.1 outlines the content and cognitive domains for mathematics. Table 2.2 outlines the content and cognitive domains for science.

Table 2.1: Mathematics content and cognitive domains and the proportion of assessment for each domain

Content domain	Topic	Per cent assessment
Number	Whole numbers Fractions and decimals Integers Ratio, proportion and percentages	30%
Algebra	Patterns Algebraic expressions Equations and formulae Relationships	25%
Measurement	Attributes and units Tools, techniques, formulae	15%
Geometry	Lines and angles Two- and three-dimensional shapes Congruence and similarity Location and spatial relationships Symmetry and transformations	15%
Data	Data collection and organiszation Data representation Data interpretation Uncertainty and probability	15%

Cognitive domain	Learner behaviour	Per cent assessment
Knowing facts and procedures	Recall Recognise/identify Compute Use tools	15%
Using concepts	Know Classify Represent Formulate Distinguish	20%
Solving routine problems	Select Model Interpret Apply Verify and check	40%
Reasoning	Hypothesise/conjecture/predict Analyse Evaluate Generalise Connect Synthesise/integrate Solve non-routine problems Justify/prove	25%

Table 2.2: Science content and cognitive domains and the proportion of assessment for each domain

Content domain	Topic	Per cent assessment
Life sciences	Types, characteristics and classification of living things Structure, function and life processes in organisms Cells and their functions Development and life cycles of organisms Reproduction and heredity; diversity, adaptation and natural selection Ecosystems and human health	30%
Chemistry	Classification and composition of matter Particulate structure of matter Properties and uses of water Acids and bases Chemical change	15%

Content domain	Topic	Per cent assessment
Physics	Physical states and changes in matter Energy types, sources and conversions Heat and temperature Light Sound and vibration Electricity and magnetism Forces and motion	25%
Earth science	Earth's structure and the physical features Earth's processes, cycles and history Earth in the solar system and the universe	15%
Environmental science	Changes in population Use and conservation of natural resources Changes in environments	15%
Cognitive domain	**Learner behaviour**	**Per cent assessment**
Factual knowledge	Recall/recognise Define Describe Use tools and procedures	30%
Conceptual understanding	Illustrate with examples Compare/contrast/classify Represent/model Relate Extract/apply information Find solutions Explain	35%
Reasoning and analysis	Analyse/interpret/solve problems Integrate/synthesis Hypothesise/predict Design/plan Collect/analyse/interpret data Draw conclusions	35%

The mathematics and science tests were developed internationally in a collaborative manner. Two different types of questions (multiple-choice questions and constructed-response questions) were included in the pool of TIMSS questions. The overriding principle in the construction of achievement tests for TIMSS 2003 was to produce assessment instruments that would generate valid and reliable data. To achieve a valid assessment of the two subjects, a substantial number of assessment items was needed. To accommodate the large number of items required in the limited testing time available (about 90 minutes per learner), TIMSS used a matrix-sampling technique. This technique involved dividing the item pool among a set of 12 learner booklets. The questions were assigned to the booklets in such a way that the combined responses of individual learners provided a comprehensive picture of the entire learner population's achievement.

Each learner completed only one of the 12 achievement booklets for the study. These were distributed equally among the class learners, in order that approximately equal numbers of learners responded to each booklet.

Background questionnaires

TIMSS included information on the educational contexts within which learners learn mathematics and science. Thus, TIMSS administered questionnaires to curriculum specialists, learners in participating schools, their mathematics and science teachers, and their school principals.

Curriculum questionnaires were designed to collect information on the organisation of each country's mathematics and science curricula, and on the subjects' content intended to be covered up to Grade 8. The national research co-ordinator in each country was responsible for the completion of these questionnaires.

Learner questionnaires were completed by each learner taking the assessment. The questionnaires focused on aspects of learners' home and school lives, classroom experiences, self-perception and attitudes about mathematics and science, homework and out-of-school activities, computer use, home educational supports, and other basic demographic information.

Mathematics teacher questionnaires and *science teacher questionnaires* were given to the maths and science teachers in the TIMSS classes. These questionnaires provided information on the teachers' backgrounds, beliefs, attitudes, educational preparations, and teaching loads, as well as the pedagogic approach adopted in the classroom. The questionnaires examined characteristics of the classes tested in TIMSS: instructional time, materials, activities for teaching mathematics and science, promoting learners' interest in the subject, assessment practices, and home-school connections. Included in the *learner* and *teacher questionnaires* were additional questions specific to South African learners and teachers.

School questionnaires were answered by the school principals. These asked about enrolment and staffing, resources available to support mathematics and science instruction, school goals, the role of the school principal, instructional time, and the school climate.

All questionnaires were designed to take about 30 minutes to complete.

Sampling

The TIMSS sampling design was a three-stage stratified cluster design (*TIMSS 2003 School Sampling Manual*), and involved:
- Selecting a sample of schools from all eligible schools;
- Randomly selecting a mathematics and science class from each sampled school; and
- Sampling learners within a sampled class in cases where the number of learners in a class is greater than 40.

Selecting a sample of schools for test administration

TIMSS is a population survey and the sample of learners is representative of the population from which it is drawn – in South Africa, these are Grade 8 learners. For South Africa, the SRN database was used to select the sample of schools. Statistics Canada assisted in selecting the random sample. While TIMSS recommended a minimum sample size of 150 schools and 4 500 learners, South Africa oversampled the number of schools and learners in order to generate provincial statistics. The sample size was 265 schools and approximately 9 000 learners across the nine provinces.

The sample was explicitly stratified by two dimensions:

- By province; and
- By the language of teaching and learning (English and Afrikaans were the only languages of instruction chosen by schools and indicated in the SRN 2000.)

It was anticipated that a 100 per cent participation rate of selected schools might not be possible. To maximise the number of schools participating, a first- and second-replacement school, displaying similar characteristics to the originally sampled study school, was also selected.

Table 2.3 indicates the number of schools sampled, and the number of schools and learners that participated in TIMSS 2003 in South Africa. Of the 265 schools sampled, ten schools withdrew at a late stage and 14 were replacement schools. Appendix 1 of this report provides a GIS plot of TIMSS 2003 participating schools.

Table 2.3: TIMSS Grade 8 schools sampled, schools in which instruments were administered, and number of learners

Province	Schools sampled	Sampled schools participating	First replacement school	Second replacement school	Total schools	Total learners
KwaZulu-Natal	48	43	2	1	46	1 632
Eastern Cape	33	29	3	1	33	943
Western Cape	25	22	1		23	813
Limpopo	32	31			31	1 145
Free State	25	24	1		25	867
Mpumalanga	25	23	1		24	962
North West	25	25			25	946
Gauteng	27	20	3		23	774
Northern Cape	25	24	1		25	870
Total	265	241	12	2	255	8 952

Selecting a class for the administration of TIMSS

Once the schools were selected, permission was sought from the school principal and the district offices for the administration of the TIMSS instruments. The next step was to select one intact Grade 8 class. The TIMSS design involved the selection of intact classes, rather than a random sample of learners from across the grade. Having an intact class ensured that the mathematics and science teacher could be matched. Furthermore, the beliefs and practices of the teachers could provide contextual information that might help explain the achievements and attitudes of their learners. However, it must be remembered that the teachers who took part in the studies were not a representative sample; rather, they were mathematics and science teachers who taught a representative sample of learners. Furthermore, the choice of intact classes meant that it was not possible to compare across schools.

Schools submitted a list of the number of Grade 8 classes. Using a systematic probability-proportionate-to-size (PPS) technique; one class was randomly selected for test administration. The assumption was that the average class size would be 40 and that all learners in the class would participate in the study.

Sampling learners within a sampled class

All learners within a classroom were expected to take part in the TIMSS assessment. For large class sizes (over 50 learners), we sub-sampled from the whole class for a group of 40 learners. This was done using a systematic sampling method, whereby all learners in a sampled classroom were assigned equal selection probabilities.

Field testing of TIMSS achievement items

After each TIMSS assessment cycle, some items are released for public use while the others are kept secure to measure trends over time. The development of new items for TIMSS 2003 started with the revision of the existing frameworks to reflect changes in curriculum and instruction in participating countries.

To replace assessment items that had been released, countries submitted items for review by subject specialists. The ISC conducted an item-writing training workshop for countries intending to submit items for possible inclusion in TIMSS 2003. A manual on *Item-writing Guidelines for TIMSS* was developed and distributed. The primary purpose of the manual, in conjunction with the workshop, was to provide information and advice on writing and reviewing items for the TIMSS tests.

Each country was responsible for the development of items in the subject of their choice. South Africa developed science items during the first two weeks in July 2001. The items and scoring guides from each country were reviewed and revised by the Science and Mathematics Item Review Committee, according to TIMSS criteria.

The new items were pilot-tested in 2002 in most of the participating countries. The piloting of the new items was co-ordinated and supported by the TIMSS and PIRLS ISC. South Africa tested Grade 9 learners instead of Grade 8 learners, since the TIMSS 2003 field test took part during the first half of the academic year. Intact mathematics classes were sampled and all learners in the mathematics class also took science as well. The mathematics and science teachers of these classes were asked to complete the teacher questionnaires. The sample for the field test comprised 25 schools from Gauteng.

Results from the pilot test from each country were pooled and used to evaluate item difficulty, how well items discriminated between high- and low-performing learners, the effectiveness of distracters in multiple-choice items, scoring suitability and reliability for constructed-response items, and evidence of bias towards or against participating countries or in favour of boys or girls. The suitability of the item was determined for the international sample, rather than for specific countries.

Main administration of TIMSS

Each country collected its own data. Procedures for data collection are outlined in *TIMSS Survey Operations Manuals*. These manuals have been designed to ensure that high-quality, internationally comparable data will be available for analysis. Training was also provided by the international project team to explain survey operations. If a country deviated from the prescribed procedures without prior approval, it ran the risk of losing the ability to measure trends properly, or to compare data with other countries participating in the study. TIMSS instruments were administered at the end of the Grade 8 academic year – for southern-hemisphere countries the TIMSS administration was conducted between October and November 2002, and for the northern-hemisphere countries the administration took place between March and May 2003.

Test administration in South Africa was carried out from 21 October to 1 November 2002. Schools were telephoned and appointments set up. School staff were used to assist with the class lists and logistical arrangements, such as identifying and preparing testing locations. An outside agency (AC Nielsen and Mictert) was chosen to administer the instruments in schools. Each of the data-collection agencies had their teams of data collectors. At the end of the data-collection process, all instruments and questionnaires were returned to the HSRC.

Training of data collectors

The HSRC trained the data collectors on administering TIMSS in the schools. The TIMSS ISC had prepared a manual for the training of data collectors.

Quality assurance of fieldwork

Each country was responsible for conducting quality-control procedures. When data was collected, a team of HSRC researchers sampled about 15 per cent of the schools for quality assurance. In addition, there was a quality control observer appointed by the ISC, who visited ten schools. The ISC also developed manuals for the monitors.

Scoring of constructed responses

The open-ended items were scored and coded. A scoring guide was developed for every open-ended item included in the TIMSS assessment. HSRC researchers were responsible for co-ordinating and monitoring the scoring and for coding the open-ended items. One HSRC researcher was trained by the TIMSS & PIRLS ISC in the coding system employed by TIMSS. A group of mathematics and science teachers was trained in data scoring – according to the data-scoring guidelines provided by the ISC. These teachers were responsible (in November and December 2002) for coding and scoring the open-ended items. To gather within-country agreement among coders, systematic sub-samples of the responses of at least 100 learners to each constructed response items were coded independently by two coders. Information on scoring and coding reliability was documented by the ISC and used to calculate cross-country agreement among coders.

To measure trends over time, TIMSS included items from 1995 and 1999 in TIMSS 2003. To ensure that constructed-response items used in 1999 were scored in the same way in 2003, participating countries sent scored booklets from the 1999 data–collection exercise to the IEA Data Processing Center (DPC), where they were scanned and the data stored. The 2003 data scorers re-scored the TIMSS 1999 responses. Scores allocated by 2003 scorers were compared with those from 1999 to check scoring consistency during this period.

Data capture and cleaning

Data capture

The IEA DPC in Hamburg, Germany (a partner with Boston College on TIMSS), provided an integrated computer programme for data entry and data verification, known as Data Entry Manager (WinDEM). WinDEM included a series of checks and verification procedures that helped ensure the quality of the data as it was being entered. An outside agency (AC Nielsen) was contracted to capture the data. The agency staff were trained in using the software. During the data-capturing process, 25 per cent of the data was recaptured for verification purposes. The data-capturing agency had to provide data with an error rate less that 0.1 per cent after verification.

Data cleaning

The data underwent a rigorous cleaning process at the HSRC, using software supplied by the DPC. The cleaning process included the following:

- Document and structure check;
- Identification variable cleaning;
- Linkage check; and
- Resolving inconsistencies in background questionnaire data.

When all the data passed the WinDEM quality-control checks, it was dispatched to the DPC for further checking and processing.

Data processing

TIMSS reported trends in learner achievement in both the general areas of mathematics and science and the major content areas related to each subject. As each learner responded to only part of the assessment, TIMSS relies primarily on item response theory (IRT) scaling methods to provide estimates of what learner achievement would be if they had responded to all the test items.

Scaling

The IRT scaling uses the multiple imputations, or 'plausible values', method to obtain proficiency scores in mathematics and science and their content areas for all learners, even though each learner responded to only a part of the assessment item pool (Yamamoto & Kuick 2000). IRT analysis provided a common scale on which performance could be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores provide estimates of how learners within countries vary and information on percentiles of performance.

Estimation

To improve reliability, TIMSS scaling methodology drew upon information gained on learners' background characteristics, as well as their responses to the achievement items. This approach, known as 'conditioning', enables reliable scores to be produced, even though individual learners responded to relatively small subsets of the total mathematics or science pool. Rather than estimating learner scores directly, TIMSS combined information about item characteristics, learner responses to the items that they took and learner background information, to estimate learner achievement distributions. Having determined the overall achievement distribution, TIMSS estimated each learner's achievement, conditional on the learner's responses to the items that they took and the learner's background characteristics. To account for error in this imputation process, TIMSS drew five such estimates, or 'plausible values', for each learner on the scales.

Reporting TIMSS achievement scores

The TIMSS scale average over the countries was set at 500, and the standard deviation at 100. In addition to scales for mathematics and science overall, TIMSS created IRT scales for each of the content domains for the 2003 data.

The *TIMSS 2003 International Mathematics Report* (Mullis et al. 2004) and the *TIMSS 2003 International Science Report* (Martin, Mullis, Gonzales & Chrostowski, 2004) summarised learners' mathematics and science achievement in each participating country. TIMSS 2003 also collected information about the homes, schools, classrooms and teachers of the participating learners, as well as the mathematics and science curriculum in each country. The TIMSS 2003 international mathematics and science reports summarised much of this information, combining data into composite indices and showing an association with achievement, where appropriate. These two international reports form the basis of the South African TIMSS Report.