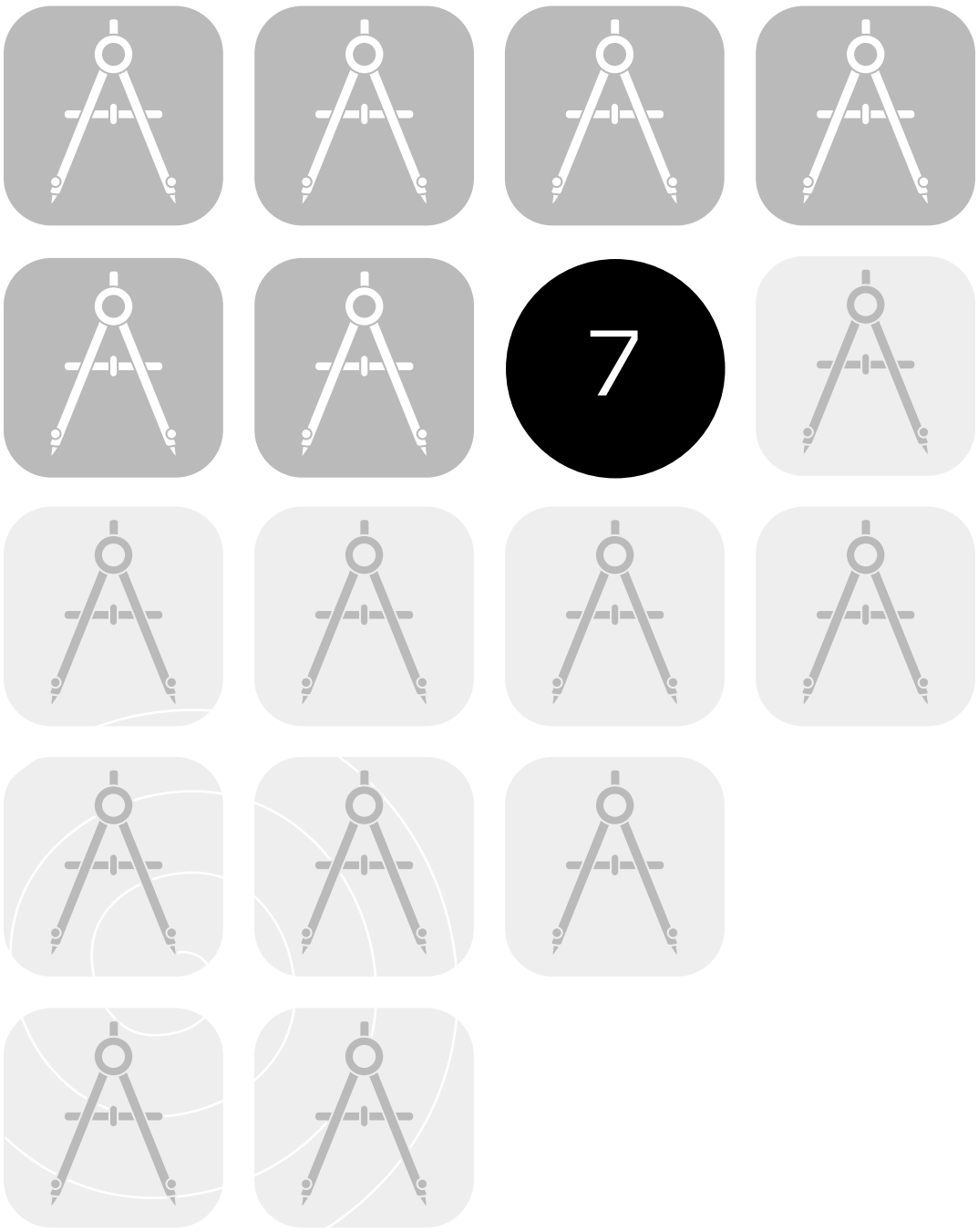
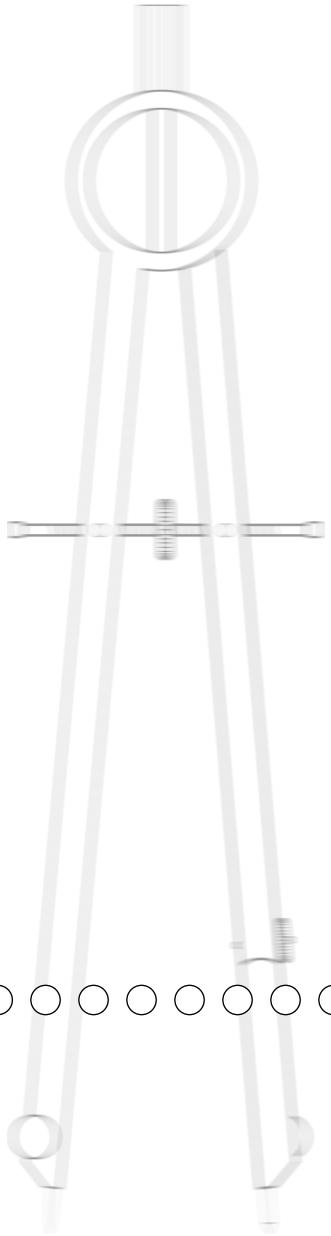


TIMSS Field Operations and Data Preparation

Eugenio J. Gonzalez
Dirk Hastedt



7 TIMSS Field Operations and Data Preparation

Eugenio J. Gonzalez
Dirk Hastedt

7.1 Overview

The TIMSS 1999 data collection in each country was a very demanding exercise, requiring close cooperation between the National Research Coordinator (NRC) and school personnel - principals and teachers - and students. The first part of this chapter describes the field operations necessary to collect the data, including the responsibilities of the NRC, the procedure for sampling classrooms within schools and tracking students and teachers, and the steps involved in administering the achievement tests and background questionnaires. The second part describes the activities involved in preparing the data files at the national center, particularly the procedures for scoring the free-response items, creating and checking data files for achievement test and questionnaire responses, and dispatching the completed files to the IEA Data Processing Center in Hamburg.

7.2 TIMSS 1999 Field Operations

The TIMSS 1999 field operations were designed by the International Study Center at Boston College, the IEA Data Processing Center, and Statistics Canada. They were based on procedures used successfully in TIMSS 1995 and other IEA studies, and refined on the basis of experience with the TIMSS 1999 field test.

7.2.1 Responsibilities of the National Research Coordinator

In each country, the National Research Coordinator was the key person in conducting the field operations. The NRC was responsible for collecting the data for the TIMSS assessment according to internationally agreed procedures and preparing the data according to international specifications. Earlier chapters of this report have outlined the tasks of the NRC with regard to choosing a sample of schools and translating the achievement tests and questionnaires.[1] This section focuses on NRC activities with regard to administering the assessment in participating schools.

○○○

1. See Chapter 2 for information about sampling schools, and Chapter 5 for details of the translation task.

Specifically it describes the procedures for sampling classes within schools, for tracking classes, teachers, and students in the sampled schools, and for organizing the administration of the achievement tests and questionnaires.

7.2.2 Documentation and Software

NRCs were provided with a comprehensive set of procedural manuals detailing all aspects of the data collection.

- The *Survey Operations Manual* (TIMSS, 1997a) was the essential handbook of the National Research Coordinator, and described in detail all of the activities and responsibilities of the NRC, from the moment the TIMSS instruments arrived at the national center to the moment the cleaned data files and accompanying documentation were submitted to the IEA Data Processing Center.
- The *TIMSS-R School Sampling Manual* (TIMSS, 1997b) defined the TIMSS 1999 target population and sampling goals and described the procedures for the sampling of schools.
- The *School Coordinator Manual* (TIMSS, 1997c) described the activities of the school coordinator (the person in the school responsible for organizing the TIMSS test administration), from the time the testing materials arrived at the school to the time the completed materials were returned to the national TIMSS center.
- The *Test Administrator Manual* (TIMSS, 1997d) described in detail the procedures for administering the TIMSS tests and questionnaires, from the beginning of the test administration to the return of the testing materials to the school coordinator.
- The *Scoring Guides for Mathematics and Science Free-Response Items* (TIMSS, 1998a) contained instructions for scoring the short-answer and extended-response test items.
- The *Manual for Entering the TIMSS-R Data* (TIMSS, 1998b) provided the NRC with instructions for coding, entering, and verifying the data. The manual included the codebook, which defined the variables and file formats in the data files.
- The *Manual for National Quality Control Observers* (TIMSS, 1998c) provided instructions for conducting classroom observations in a sample of participating schools.

Additionally, two software packages were supplied by the IEA Data Processing Center to assist NRCs in the main study

- The within-school sampling software (W3S), a computer program designed to help NRCs select the within-school sample, prepare the survey tracking forms, and assign test booklets to students was supplied along with its corresponding manual.
- The DATAENTRYMANAGER, a computer program for data entry and data verification was supplied along with its corresponding manual.

In addition to the manuals and software, NRCs received hands-on training in the procedures and use of the software from staff of the International Study Center, the IEA Data Processing Center, and Statistics Canada.

7.2.3 Within-School Sampling Procedures

The study design anticipated relational analyses between student achievement and teacher-level data at the class level. For field operations, this meant that intact classes had to be sampled, and that for each sampled class the mathematics and science teachers had to be tracked and linked to their students. Although intact classes were the unit to be sampled in each school, the ultimate goal was a nationally representative sample of students. Consequently, in each country a classroom organization had to be chosen that ensured that every student in the school was in one class or another, and that no student was in more than one class. Such an organization is necessary for a random sample of classes to result in a representative sample of students. In most countries at the eighth grade, mathematics classes serve this purpose well, and so were chosen as the sampling units. In countries where students attended different classes for mathematics and science, classrooms were defined on the basis of mathematics instruction for sampling purposes.[2]

The TIMSS design required that for each student in each sampled class, all eighth-grade mathematics and science teachers be identified and asked to complete a teacher questionnaire.

○○○

2. For countries where a suitable configuration of classes for sampling purposes could not be identified, TIMSS also provided a procedure for sampling individual students directly from the eighth grade.

When sampling mathematics classes in a school, the procedure was as follows

- The NRC asked the school coordinator for a list of all mathematics classes in the target (eighth) grade along with the names of their mathematics teachers.
- The school coordinator sent the requested list to the NRC.
- The NRC transcribed the information onto a document known as a *Class Sampling Form* and applied a prescribed sampling algorithm to select one or more classes.
- For each sampled class, the NRC prepared a *Teacher-Student Linkage Form* designed to link the students in the class to each of their eighth-grade mathematics and science teachers. The form was then sent to the school coordinator to be completed.
- The school coordinator completed the Teacher-Student Linkage Form by listing all of the students in the class (name or identification number, date of birth, and sex), together with their mathematics and science teachers and classroom identifiers as necessary, and returned it to the NRC.
- From the information provided in the Teacher-Student Linkage Form, the NRC produced a *Student Tracking Form*, which listed all students in the class to be tested with their TIMSS identification numbers and booklet assignments, and a *Teacher Tracking Form,* which listed all mathematics and science teachers of the students in the class, their student-teacher link numbers, and their questionnaire assignments. These forms were sent to the school coordinator along with the test instruments.
- During the test administration, the test administrator and school coordinator used the tracking forms to record student and teacher participation, and returned them to the NRC after the test administration together with the completed test booklets and questionnaires.

7.2.4 Excluding Students from Testing

Although all students enrolled in the target grade were part of the target population and were eligible to be selected for testing, TIMSS recognized that some students in every school would be unable to take part in the 1999 assessment because of some physical or mental disability. Accordingly, the sampling procedures provide for the exclusion of students with any of several disabilities (see Chapter 2). Countries were required to track and

account for all excluded students, and were cautioned that excluding an excessive proportion would lead to their results being annotated in international reports. It was important that the conditions under which students could be excluded be carefully delineated, because the definition of "disabled" students varied considerably from country to country.

7.2.5 Survey Tracking Forms

As is evident from the description of the within-school sampling procedure provided earlier, TIMSS 1999 relied on a series of "tracking forms" to implement and record the sampling of classes, teachers, and students. It was essential that the tracking forms were completed accurately, since they made explicit exactly who should be given which instruments, and recorded what happened in each school. In addition to facilitating the data collection, the tracking forms provided essential information for the computation of sampling weights and for evaluating the quality of the sampling procedures All tracking forms were retained for review by staff of the International Study Center.

Survey tracking forms were provided for sampling classes and students; for tracking schools, classes, teachers, and students; for linking students and teachers; and for recording information during test administration. Each of these forms is described below.

7.2.6 Linking Students, Teachers, and Classes

Within each school, a class identification number (ID) was assigned to each class in the target grades listed on the class tracking form. The class ID consisted of the three-digit school ID plus a two-digit identification number for the class within the school.

Each student listed on the student tracking form was assigned a student identification number. This was a seven-digit number consisting of the five-digit class ID plus a two-digit number corresponding to the student's sequential position in the student tracking form. All students listed on the student tracking form, including those marked for exclusion, had to be assigned a student ID.

All mathematics and science teachers of the selected classes (those listed on the teacher tracking form) were assigned a teacher ID that consisted of the three-digit school ID plus a two-digit number of the teacher within the school. Since a teacher could be teaching both mathematics and science to some or all of

the students in a class, it was necessary to have a unique identification number for each teacher/class and teacher/subject combination. This was achieved by adding a two-digit link number to the five digits of the teacher ID, giving a unique seven-digit teacher/class identification number. Careful implementation of these procedures was necessary so that later each class could be linked to a teacher, and student outcomes could be analyzed in relation to teacher-level variables.

7.2.7 Assigning Testing Materials to Students and Teachers

Eight different test booklets were distributed to the students in each sampled class. Each student was required to complete one booklet, and the student questionnaire. Booklets were assigned to students by the NRC using a random assignment procedure, after which the assignment was recorded on the student tracking form.

Each teacher listed on the teacher tracking form was assigned a mathematics or a science teacher questionnaire. Where teachers taught both mathematics and science to the class, every effort was made to collect information about both. However, NRCs had the final decision as to how much response burden to place on such teachers.

7.2.8 Administering the Test Booklets and Questionnaires

The school coordinator was the person in the school responsible for organizing the administration of the TIMSS 1999 tests. This could be the principal, the principal's designee, or an outsider appointed by the NRC with the approval of the principal. The NRC was responsible for ensuring that the school coordinators were familiar with their responsibilities.

The major responsibilities of the school coordinators are detailed in the school coordinator manual (TIMSS, 1997c). Prior to the test administration the tasks for the school coordinator included:

- Providing the NRC with all information necessary to complete the various tracking forms
- Checking the testing materials when they arrived in the school to ensure that everything was in order
- Ensuring that the testing materials were kept in a secure place before and after the test administration
- Arranging the dates of the test administration with the national center

- Arranging for a test administrator and giving a briefing on the TIMSS 1999 study, the testing materials, and the testing sessions
- Working with the school principal, the test administrator, and the teachers to plan the testing day; this involved arranging rooms, times, classes and materials

The Test Administrator was responsible for administering the TIMSS tests and student questionnaires. Specific responsibilities were described in the test administrator manual (TIMSS, 1997d), and included:

- Ensuring that each student received the correct testing materials which were specially prepared for him or her
- Administering the test in accordance with the instructions in the manual
- Ensuring the correct timing of the testing sessions by using a stopwatch and recording the time when the various sessions started and ended on the test administration form
- Recording student participation on the student tracking form.

The responsibilities of the school coordinator after the test administration included:

- Ensuring that the test administrator returned all testing materials, including the completed student tracking form, the test administration form, and any unused materials
- Calculating the student response rate and arranging for makeup sessions if it was below 90%
- Distributing the teacher questionnaires to the teachers listed on the teacher tracking form, ensuring that the questionnaires were returned completed, and recording teacher participation information on the teacher tracking form
- Preparing a report for the NRC about the test administration in the school
- Returning both completed and unused test materials and all tracking forms to the NRC

The NRC prepared two packages for each sampled class. One contained the test booklets for all students listed on the student tracking form and the other the student questionnaires. For each participating school, the test booklets and student questionnaires

were bundled together with the teacher tracking form and teacher questionnaires, the school questionnaire, and the materials prepared for briefing school coordinators and test administrators, and were sent to the school coordinator. A set of labels and prepaid envelopes addressed to the NRC was included to facilitate the return of testing materials.

7.3 National Quality Control Program

The International Study Center implemented an international quality control program whereby international quality control monitors visited a sample of 15 schools in each country and observed the test administration. In addition, NRCs were expected to organize a national quality control program, based upon the international model. This national program required Quality Control Observers to document data collection activities in their country. They visited a 10% sample of TIMSS 1999 schools, observed actual testing sessions, and recorded compliance of the test administration with prescribed procedures.

The International Study Center prepared *The Manual for National Quality Control Observers* (TIMSS, 1998c) which contained information about TIMSS 1999, and detailed the role and responsibilities of the National Quality Control Observers.

7.4 Data Preparation

In the period immediately following the administration of the TIMSS 1999 tests, the major tasks for the NRC included retrieving the materials from the schools; recruiting and training scorers to score the free-response items; scoring these items, including double scoring a 25% reliability sample; entering the data from the achievement tests and background questionnaires; submitting the data files and materials to the IEA Data Processing Center; and preparing a report on survey activities.

When the testing materials were received back from the schools, NRCs were to do the following:

- Check that the appropriate testing materials were received for every student listed on the student tracking form
- Verify all identification numbers on all instruments that were not precoded at the national center
- Check that the participation status recorded on the tracking forms matched the information on the test instruments

- Follow up on schools that did not return the testing materials or for which forms were missing, incomplete, or inconsistent

NRCs then organized the tests for scoring and data entry. The procedures involved were designed to maintain identification information that linked students to schools and teachers, minimize the time and effort spent handling the booklets, ensure reliability in the free-response coding, and document the reliability of the coding.

7.4.1 Scoring the Free-Response Items

Reliable application of the scoring guides to the free-response questions, and empirical documentation of the reliability of the scoring process, were critical to the success of TIMSS 1999. The survey operations manual (TIMSS, 1997a) contained information about arranging for staff and facilities for the free-response scoring effort required for the TIMSS 1999 main survey; for effective training of the scorers; and for distributing booklets to scorers to accomplish the scoring for the main data set. Countries were to double score a 25% sample to document scoring reliability.

For most countries, the scope of the free-response scoring effort was substantial. The main survey contained 68 free-response questions. Each of the 8 booklets had between 9 and 14 free-response questions. On average, each country had to score about 50,000 student responses.

To ascertain the staff requirements for free-response scoring, it was necessary to estimate the amount of scoring to be done and the amount of time available to do it, and also to make provision for staff training and for clerical and quality control throughout the operation. The International Study Center recommended at least one half-day of training on each of the 8 booklets, for a total of about a week for training activities.

In scoring the free-response items it was vital that scoring staff apply the scoring rules consistently and in the same way in all participating countries. Hence, in selecting those who were to do the scoring, NRCs took care to arrange for persons who were conscientious and attentive to detail, knowledgeable in mathematics and science, and willing to apply the scoring guides as stated, even if they disagreed with a particular definition or category. Preference was given to individuals with educational backgrounds in the math-

ematics and science curriculum areas or who had taught at the middle school level. Good candidates for scoring included teachers, retired teachers, college or graduate students, and staff of education agencies or ministries and research centers.

7.4.2 Preparing Materials to Train the Scorers

The success of assessments containing free-response questions depends upon the reliability of scoring student responses. In TIMSS 1999, reliability was assured through the provision of scoring guides (manuals), extensive training in their use, and monitoring of the quality of the work. In addition, TIMSS 1999 provided training packets for training in selected questions, and practice papers to help scorers achieve a consistent level of scoring.

Each scorer received a copy of the *TIMSS 1999 Main Survey Scoring Guides for Mathematics and Science Free-Response Items* (TIMSS, 1998a). This document explained the TIMSS scoring system, which was designed to produce a rich and varied profile of the range of students' competencies in mathematics and science.[3]

At the international scoring training meetings, NRCs received training packets containing example responses and practice papers to help them achieve accuracy and consistency in scoring. For scoring guides that were difficult, example responses were selected to illustrate the scoring categories. The scores on these responses were explained and attached to the scoring guides. Practice sets were created for the more difficult guides. These papers illustrated a range of responses, beginning with several clear-cut examples. About 10 to 15 responses were enough for most guides, but sometimes more practice was necessary.

7.4.3 Documenting the Reliability of the Free-Response Scoring

In order to demonstrate the quality of the TIMSS 1999 data, it was important to document the agreement between scorers. To establish the scoring reliability, NRCs were required to have a 25% random sample of each booklet type independently scored by two scorers. The degree of agreement between the two scores assigned was a measure of the reliability of the scoring process. The two scorers did not know the scores assigned by the others.

○○○

3. The TIMSS scoring scheme for free-response items is described in Chapter 3

Since the purpose of the double scoring was to document the consistency of scoring, the procedure used for scoring the booklets in the reliability sample had to be as close as possible to that used for scoring the booklets in general. The recommended procedure was designed to blend the scoring of the sample in with the normal scoring activity, to take place throughout the scoring process, and to be systematically implemented across student responses and scorers.

7.4.4 Scoring the Free-Response Items

TIMSS 1999 recommended that scorers be organized into teams of about six, headed by a team leader. The leader's primary responsibility was to monitor scoring reliability by continually checking and rechecking the scores that scorers had assigned. This process, known as back-reading, was essential for identifying scorers who did not understand particular guides or categories. Early detection of any misunderstandings permitted clarification and rectification of mistakes before too many responses had been scored. The back-reading systematically covered the daily work of each scorer. If a particular scorer appeared to have difficulty, however, then the percentage of back-reading for that scorer was increased. Any errors discovered were brought to the attention of the scorer responsible and corrected immediately. If a scorer was found to have been consistently making an error, then all of the booklets scored by that person were checked and any errors corrected.

In scoring the booklets for the main data set, scorers entered their scores directly into the student booklets. Therefore, in order that the reliability scoring be done "blind" (i.e., so that the two scorers did not know each other's scores), the reliability scoring had to be done before the scoring for the main data, and the reliability scores had to be recorded on a separate scoring sheet, and not in the booklets.

To implement the scoring plan effectively it was necessary that the scorers be divided between two equivalent teams (Team A and Team B), and that booklets be divided into two equivalent sets (Set A and Set B). The scorers in Team A scored 25% of the booklets in Set B and all the booklets in Set A, while the scorers in Team B scored 25% of the booklets in Set A and all of the booklets in Set B. Each team, therefore, handled both sets of booklets. For the set it handled first, the team scored every fourth booklet and recorded the results on a separate answer sheet (this was the reliability sample). In the other set, the team scored all booklets and wrote the scores directly into the booklets.

Periodically during the day, the Team B scorers scored the reliability sample (every fourth booklet) in the Set A batches, while the Team A scorers scored the reliability sample in the Set B batches. It was important that every fourth booklet be scored, and not just the top quarter in the set. When the reliability scoring was finished, Team B scorers marked it as completed and forwarded the batch to the Team A scorers. Similarly, the Team A scorers forwarded their scored reliability booklets from Set B to the Team B scorers. Once the booklets from Set A had been distributed to Team A scorers and the Set B booklets to the Team B scorers, all the free-response items were scored, and the scores were entered directly into the booklets.

7.5 Data Entry

The DPC provided an integrated computer program for data entry and data verification known as the DATAENTRYMANAGER (DEM). This program worked on all IBM-compatible personal computers running under DOS, OS/2 or Windows 3.x, 95 or 98. It facilitated data entry directly from the tracking forms and test instruments and provided a convenient checking and editing mechanism. DEM also offered data and file management capabilities, interactive error detection, reporting, and quality control procedures. Detailed information and operational instructions were provided in the DATAENTRYMANAGER manual. Since DEM incorporated the international codebooks describing all variables, use of the software ensured that the data files were produced according to the TIMSS 1999 rules and standards for data entry.

Although use of DEM for all data entry tasks was strongly recommended, NRCs were permitted to use their own procedures and computer programs, as long as all data files conformed to the specifications of the international codebooks. NRCs who chose not to use DEM were responsible for ensuring that all data files were delivered to the DPC in the international format.

Even if NRCs did not use the DEM program for data entry, they still had to apply the data verification options of this program to verify their data before sending them to the DPC. The DEM data-checking facility could: (i) identify a range of problems in the identification variables, and invalid codes; and (ii) identify problems in the structure of the data files, which could then be fixed before submission to the NRC.

Data files were regarded as having been satisfactorily checked only if the reports generated by the DEM program indicated no errors.

During the TIMSS 1999 main survey operations, data were gathered from several sources, including students, teachers, and principals, as well as from a range of tracking forms. Before beginning data entry, the NRC needed to ensure that the corresponding tracking forms and instruments had been completed and sorted correctly. The data were entered into one of six data files, as follows:

- The **school background file** contained information from the school background questionnaire
- The **mathematics teacher background file** contained information from the mathematics teacher questionnaire
- The **science teacher background file** contained information from the science teacher questionnaire
- The **student background file** contained data from the student background questionnaire
- The **student achievement file** contained the achievement test booklet data
- The **free-response scoring reliability file** contained the reliability data from the scoring of the free-response items

When all data files had passed the DEM quality control checks, they were dispatched to the IEA Data Processing Center in Hamburg for further checking and processing.

7.5.1 Survey Activities Report

NRCs were requested to maintain a record of their experiences during the TIMSS 1999 data collection and to send a report to the International Study Center when data-collection activities were completed. This should describe any problems or unusual occurrences in selecting the sample or securing school participation, translating or preparing the data-collection instruments, administering the tests and questionnaires in the schools, scoring the free-response items, or creating and checking the data files.

7.6 Summary

This chapter has summarized the design and implementation of the TIMSS 1999 field operations from the point of first contact with the sampled schools to the return of the cleaned data files to the IEA Data Processing Center. Although the procedures were sometimes complex, each step was clearly documented in the TIMSS operations manuals and supported by training sessions at the NRC meetings. NRC reports indicated that the field operations went well, and that the TIMSS data were of high quality.

References

TIMSS (1997a). *Survey Operations Manual* (TIMSS 1999 Doc. Ref. No. 98-0026). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1997b). *TIMSS-R School Sampling Manual* (TIMSS 1999 Doc. Ref. 97-0012). Prepared by Pierre Foy, Statistics Canada. Chestnut Hill, MA: Boston College.

TIMSS (1997c). *School Coordinators Manual* (TIMSS 1999 Doc. Ref. 98-0024). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1997d). *Test Administrator Manual* (TIMSS 1999 Doc. Ref. 98-0025). Prepared by the IEA Data Processing Center. Chestnut Hill, MA: Boston College.

TIMSS (1998a). *Scoring Guides for Mathematics and Science Free-Response Items* (TIMSS 1999 Doc. Ref. 98-0049). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998b). *Manual for Entering the TIMSS-R Data* (TIMSS 1999 Doc. Ref. 98-0028). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

TIMSS (1998c). *Manual for National Quality Control Observers* (TIMSS 1999 Doc. Ref. 98-0044). Prepared joined by the International Study Center at Boston College and the IEA's Data Processing Center. Chestnut Hill, MA: Boston College.