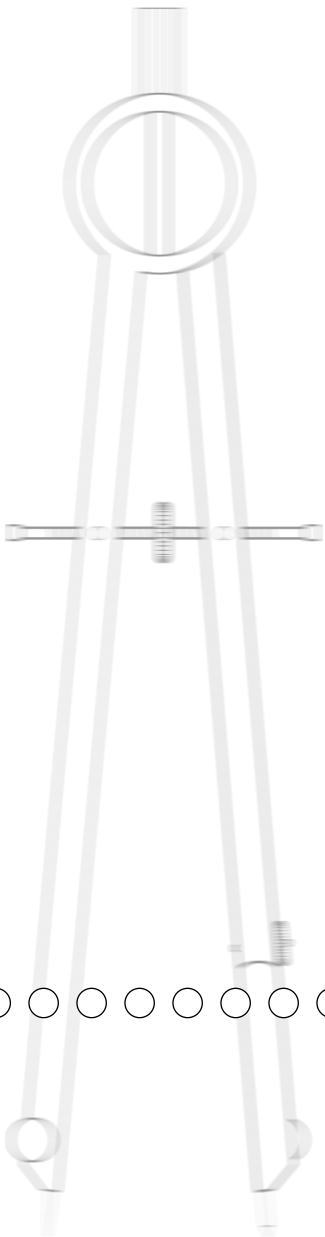
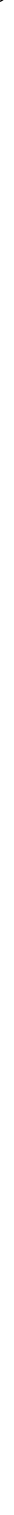
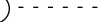


## Data Collection and Data Preparation for TIMSS 1999 Benchmarking

Dward Moore







# 8

## Data Collection and Data Preparation for TIMSS 1999 Benchmarking

Dward Moore

### 8.1 Overview

This chapter discusses the data collection and processing activities conducted for the TIMSS 1999 Benchmarking study. Under contract with the International Study Center at Boston College, Westat was responsible for data collection and preparation. In particular, Westat coordinated the within-school sampling and data collection activities. Westat subcontracted National Computer Systems (NCS) to process the data, produce the data collection forms, receive and score the completed assessment materials, and prepare the database for final analysis.

### 8.2 Field Operations and Data Collection

Data collection for the TIMSS 1999 Benchmarking study occurred March 15 through June, 1999, concurrently with data collection for the U.S. national sample for TIMSS 1999 and for most of the TIMSS 1999 Northern Hemisphere countries. A number of people were involved in the data collection effort in the U.S., including the Westat field manager, jurisdiction coordinators, field supervisors, school coordinators, and test administrators. These individuals assisted in gaining access, gathering necessary sampling information, and scheduling and administering the assessment in the sampled schools.

#### 8.2.1 Field Manager

The TIMSS 1999 Benchmarking data collection activities were directed by a Westat field manager who oversaw the work of 20 field supervisors. The field manager also visited sites to assess whether test administration procedures were implemented correctly and uniformly.

#### 8.2.2 Jurisdiction Coordinators

Each participating jurisdiction selected a contact person to work with the Westat field staff in gaining participation of the sampled schools and to act as the conduit in collecting the necessary information from the schools.

### 8.2.3 Field Supervisors

The field supervisors were mainly responsible for coordinating data collection activities in the schools assigned to them within a geographic area. They attended a training session and received a manual developed by Westat. Field supervisors contacted school coordinators to confirm schedules, ascertain that the testing materials had arrived at the schools, and finalize assessment arrangements. They also trained the test administrators. Westat provided a script for the field supervisors to use in the training sessions.

### 8.2.4 School Coordinators

Each participating school appointed a school-level coordinator, usually a teacher or principal, to be responsible for obtaining information on sampling classes within a school and making preparations for the assessment: scheduling test administration, receiving testing materials, distributing the Teacher Questionnaires, and completing (or selecting a designee who would complete) the School Questionnaire.

### 8.2.5 Test Administrators

Test administrators were responsible for preparing for test administration and for going to the schools on the agreed-upon assessment date to collect the Teacher and School Questionnaires and to administer the assessments. Test administrators were expected to spend up to four hours studying the *TIMSS 1999 Test Administrator Manual* (TIMSS, 1998) and attending the training session run by the field supervisor before the assessment day. Adherence to the standard procedures set forth in the *TIMSS 1999 Test Administrator Manual* was emphasized.

## 8.3 Benchmarking Manuals

Westat followed the data collection procedures detailed in the international manuals (see chapter 7 for a list of manuals provided to NRCs). Based on these materials, Westat developed a guide for field supervisors during data collection. In addition, Westat adapted the *TIMSS 1999 Test Administrator Manual* (TIMSS, 1998) to apply to the Benchmarking context.

### 8.3.1 Supervisors Manual

Westat developed the *TIMSS 1999 Benchmarking Supervisors Manual* (Westat, 1998a) to guide supervisors in conducting within-school sampling; recruiting, hiring and training test administrators; contacting schools to secure participation and scheduling; and observing at least one testing session conducted by each test administrator.

### 8.3.2 Benchmarking Test Administrator Manual

The *TIMSS 1999 Benchmarking Test Administrators Manual* (Westat, 1998b) detailed the standard international procedures and included additional information to help test administrators conduct a high-quality session. The manual included an overview of the Benchmarking study, a section on classroom management, question-by-question specifications for the Student Questionnaire, and administrative procedures such as completing the expense reports.

## 8.4 Within-School Sampling

Once jurisdiction coordinators secured cooperation of the sampled schools, each school coordinator obtained the information necessary for sampling classes within the school. Coordinators followed the standard international procedure for sampling classes and teachers. The process and forms used were as follows:

- The school coordinator sent a list of the eighth-grade mathematics classes and their teachers' names to the jurisdiction coordinator using the Class Listing Form. This form, developed by Westat, included instructions and definitions regarding the classes to be listed.
- The jurisdiction coordinator forwarded the class listing to Westat staff who selected the sample of two intact mathematics classrooms using the Class Sampling Form. The Class Sampling Form used for Benchmarking was identical to the main survey form used in the United States. It differed from the international version in that it specified minimum class size rather than minimum cluster size. (Minimum class size, 10, is given as half the minimum cluster size). In the Class Sampling Forms, pseudoclasses or classes of students "not taking math" were formed as necessary following the international procedures required by the International Study Center.

- The selected mathematics classes were each specified on a Student-Teacher Worksheet. The jurisdiction coordinator mailed the Student-Teacher Worksheet to the school coordinator along with a cover letter asking for lists of the students from the selected mathematics classes (their names, gender, and birth dates) and each student's science class and science teacher's name.
- Westat staff prepared the Teacher-Student Linkage Forms, Teacher Tracking Forms, and Student Tracking Forms based on the information submitted on the Student-Teacher Worksheet. They passed the completed forms to jurisdiction coordinators who mailed them to the school for use by the Westat test administrators on assessment day. In addition, Westat sent test administrators the prepared School Questionnaire and Teacher Questionnaires with a letter asking them to distribute these so that the completed responses could be picked up by the Westat assessment staff on testing day.
- The "exclusion" code, if any, was recorded only on the Student Tracking Forms, rather than entering it on both the Linkage and Student Tracking Forms as indicated in the international procedural manual. Schools were provided with a printed copy of the international "exclusion" definitions, and decisions on excluding students were made by school staff. The Westat assessment staff recorded exclusions on the Student Tracking Forms on the assessment day before testing began.
- Since the sampling activities for the TIMSS 1999 Benchmarking study in the U.S. were not centrally organized, it was decided not to use the TIMSS sampling software.

## 8.5 Test Administration

In most schools, students were assessed in two groups (the mathematics classes that were sampled). Due to class scheduling, students in some schools were assessed as one group in a large facility such as the cafeteria or library.

Test administrators followed the standard script for administering the assessment based on internationally prescribed procedures. The responsibilities of the test administrator are detailed in the *TIMSS 1999 Benchmarking Test Administrators Manual* (Westat 1998b) and included:

- Ensuring that each student received the correct testing materials that were specifically prepared for him or her

- Administering the test in accordance with the instructions in the *TIMSS 1999 Benchmarking Test Administrator Manual* and the session script
- Ensuring the correct timing of the testing session
- Recording student participation on the Student Tracking Form
- Copying student demographic information onto booklet covers
- Recording session details on the Test Administration Form and Student Response Rate Form.

Make-up sessions were requested whenever attendance in the school's sessions was below the target 90% participation. In a few schools, it was not possible to schedule make-up sessions. As each testing session was completed, test administrators sent all assessment materials to NCS for processing.

## 8.6 Quality Control Monitoring

Site visits were made by field managers and by Westat home office staff to make sure that all procedures were being implemented correctly and uniformly. Test administrators were observed by their field supervisors in at least one assessment for quality control. In addition, about 10% of schools in each jurisdiction were randomly selected for a visit by a Quality Control Monitor from the International Study Center (see chapter 9 for details).

## 8.7 Data Processing

NCS was subcontracted by Westat to process and score the TIMSS 1999 Benchmarking data for the states and districts. NCS used the same receipt control procedures that were used to process the TIMSS 1999 national assessment materials and carried out all scoring procedures for TIMSS 1999 Benchmarking simultaneously with the TIMSS 1999 national assessment. Student papers from the TIMSS 1999 national and state/district Benchmarking samples were intermingled and scored simultaneously.

When TIMSS 1999 national and Benchmarking materials came into NCS from the field, they were checked using the same specifications and were processed and scanned on parallel tracks. Each group of materials was given a letter code (TR for national and DS for Benchmarking districts and states) and a sequential number to identify the batch and document. After conducting receipt control procedures, NCS scanned the test booklets and sorted the open-ended items by item in preparation for scoring.

## 8.8 Data Entry

The data on the scannable documents were collected using NCS optical-scanning equipment that also captured images of the constructed-response items and intelligent character recognition (ICR) fields. The School and Teacher Questionnaires were entered into a data file using key-entry methods. A second person keyed the same data into a verification file. Both sets of data were then programmatically compared and discrepancies were corrected. The data were run through the NCS Pre-Edit Program and edited according to specifications that the development staff created. This ensured that the data would conform to the code-book specifications. If there were problems, the editing staff corrected them and then ran the batch through the NCS Post-Edit Program to insure that the changes were entered correctly and the data were clean.

## 8.9 Image Scoring

Because of the economy of scale due to the increased sample size and the good rate of participation by states and districts, NCS used image processing and on-line scoring for the TIMSS 1999 national and Benchmarking studies. Two of the significant advantages of this on-line system were the ease of regulating the flow of work to scorers and the ease of monitoring scoring. The system allowed for item-by-item, rather than book-by-book, scoring. Item-by-item scoring increased efficiency because scorers and trainers could focus on one item at a time, thus improving scoring reliability and validity.

On-line scoring was a valuable innovation in the TIMSS 1999 project. Training for a particular item occurs before the scoring of all responses for the item. Student papers from both the TIMSS 1999 national sample and district/state Benchmarking samples were intermingled and scored simultaneously by scorers trained in the scoring procedures for the TIMSS 1999 open-ended items. Coupled with the fact that responses were scored item by item, this meant that all occurrences of the items, whether in the national sample, in the district/state sample, or in different booklet types, were scored together following the training for that item.

## 8.10 Scoring Training

NCS prepared training materials for scorers using the international training materials distributed at the TIMSS 1999 international scoring training meeting in February 1999 as well as sample papers from the responses received from the U.S. assessment in Spring 1999. NCS scoring center staff selected a random sample of about



50 responses for each item to be scored. NCS trainers who attended the international training reviewed the responses to ensure that a wide range of responses were represented for each score point. A score was assigned to each response, and sets of papers that exemplified score points were created. Training involved:

- Presenting and discussing the item to be scored along with the item rationale
- Explaining the scoring guide to the team and discussing the anchor papers, which contain the scoring guide, the item, its scoring rationale, and the student responses that represent the various score points in the guide
- Discussing the rationale behind the scoring guide, focusing on the criteria that differentiate the levels in the guide
- Practicing scoring on a shared set of student responses
- Discussing the responses in the shared packet
- Scoring and discussing 10 to 20 practice papers, which represented the entire range of score points for that item.

After the trainer and the scoring supervisor determined that the team had reached consensus, the scoring supervisor released the work electronically to scorers through the on-line scoring system. When scorers first received live responses, they either took turns scoring these responses or worked in pairs as a final quality check before they began working on their own.

Training sets were created by May 24, 1999. The trainers received their materials June 1. Scoring began on June 14 and was completed July 9.

### 8.11 Monitoring Scoring

Trainers and scoring supervisors monitored the scoring process in three ways: (1) using a software feature that allowed for ongoing checking of scorer agreement rates; (2) using a feature that allowed for backreading the papers read by scorers; and (3) using a feature that monitored scoring rates. These software features are discussed below.

Ten percent of all of the constructed responses were scored by a second reader to obtain statistics on inter-rater reliability. The responses chosen for “second scoring” were determined systematically. Second scorers did not know that they were second scoring

or what score the first rater had assigned. NCS scoring supervisors used this information to monitor the capabilities of all scorers, to maintain uniformity of scoring, and to ensure that scorer agreement rates met minimum standards.

The inter-rater reliability feature produced on-demand reports in either of two modes—aggregate information of all first scorings versus all second scorings, or overall agreement percentage for each individual scorer. The information was displayed in a matrix format showing the instances of exact agreement, which fell along the diagonal of the matrix. Data in each cell of the matrix provided the number and percentage of cases of agreement (or disagreement). The display also contained information on the total number of second scorings and the overall percentage of cases in which two scorers agreed on a rating.

The scoring supervisor monitored each scorer's progress using the system's backreading feature. Typically, a scoring supervisor looked at about 10% of all responses graded by each scorer. With this feature, scoring supervisors could see all the papers a given rater graded, and the scores that were assigned. Where scoring supervisors disagreed with the score assigned by the first scorer, they assigned a new score, which then became the reported score.

In addition to checking rater agreement and backreading individual responses, scoring supervisors monitored workflow using a status report that displayed the number of responses scored, the number of responses scored a second time, and the number of responses remaining to be scored. This facility allowed the scoring supervisor and trainer to monitor the scoring rate accurately and to estimate the time needed to complete the scoring.

## 8.12 Completing Data Files

After open-ended scoring was completed, scores were merged with the demographic, gridded, and key-entered data. At this time, final output files were produced for each file type. The final files were checked by the Software Quality Specialists to ensure that the data adhered to the international format. In earlier editing functions, data were checked for completeness and compliance with international codebook specifications. In addition, a check was performed to verify correct linking and matching of student, teacher, and school data files. To maintain confidentiality, all personal identification in the data set was removed before the data were sent to the IEA DPC and the International Study Center at Boston College. Further checking and cleaning took place at the DPC using international procedures for data cleaning.

---

## References

---

TIMSS (1998). *TIMSS 1999 test administrator manual* (TIMSS 1999 Doc Ref. No. 98-0025). Prepared by the International Study Center at Boston College. Chestnut Hill, MA: Boston College.

Westat (1998a). *TIMSS 1999 benchmarking supervisors manual*. Prepared by Westat, Inc. Rockville, MD: Westat, Inc.

Westat (1998b). *TIMSS 1999 benchmarking test administrators manual*. Prepared by Westat, Inc. Rockville, MD: Westat, Inc.