

Heiko Sibberns
Dirk Hastedt
Michael Bruneforth
Knut Schwippert
IEA Data Processing Center

Eugenio J. Gonzalez
Boston College

The TIMSS data were processed through a closely cooperative procedure involving the TIMSS International Study Center at Boston College, the IEA Data Processing Center, the Australian Council for Educational Research, Statistics Canada, and the national research centers of the participating countries. Under the general direction of the International Study Center, each institution was responsible for specific aspects of the data processing.

The data processing consisted of six general tasks: data entry, creation of the international database, calculation of sampling weights, scaling of achievement data, analysis of the background data, and creation of the reporting tables. Although each task is crucial to ensuring the quality and accuracy of the results, data entry and the creation of the international database take center stage, since those tasks feed into the remaining four. The scaling of the TIMSS data are discussed in Chapters 7 and 8, the weighting procedures in Chapter 4, and the analysis and reporting in Chapters 9, 10, and 11. This chapter describes the process followed in data entry and the creation of the international database, and the steps that were undertaken to ensure the quality and accuracy of the international database. It also describes the responsibilities of each participant in the creation of the international database. In particular, this chapter outlines the flow of the data files between the different centers involved in the data processing; the structure of the data files submitted by each country for processing, and the resulting files that are part of the international database; the rules, methods, and procedures used for data verification and manipulation; the data products created during the data cleaning process and provided to the national centers; and the computer software used in this process.

The TIMSS international database for the primary and middle school years was released for public use in September 1997. It is available at the TIMSS website (<http://www.csteep.bc.edu/timss>) and through IEA Headquarters. The database is accompanied by a User's Guide (Gonzalez and Smith, 1997) and full documentation.

3.1 DATA FLOW

The data collected with the TIMSS survey instruments were entered into data files of a common international format at the national research centers of the participating countries. These data files were then submitted to the IEA Data Processing Center for cleaning and verification. The major responsibilities of the IEA Data Processing Center at this point were to check that the data files submitted matched the international stan-

standard and to make modifications where necessary, apply standard cleaning rules to the data to verify their consistency and accuracy, interact with the National Research Coordinators (NRCs) to ensure the accuracy of the data contained in the files, produce summary statistics of the background and achievement data for review by the TIMSS International Study Center, and, upon feedback from the individual countries and the TIMSS International Study Center, construct the international database. The IEA Data Processing Center also had primary responsibility for making all modifications to the data files and for distributing the national data files to each of the participating countries.

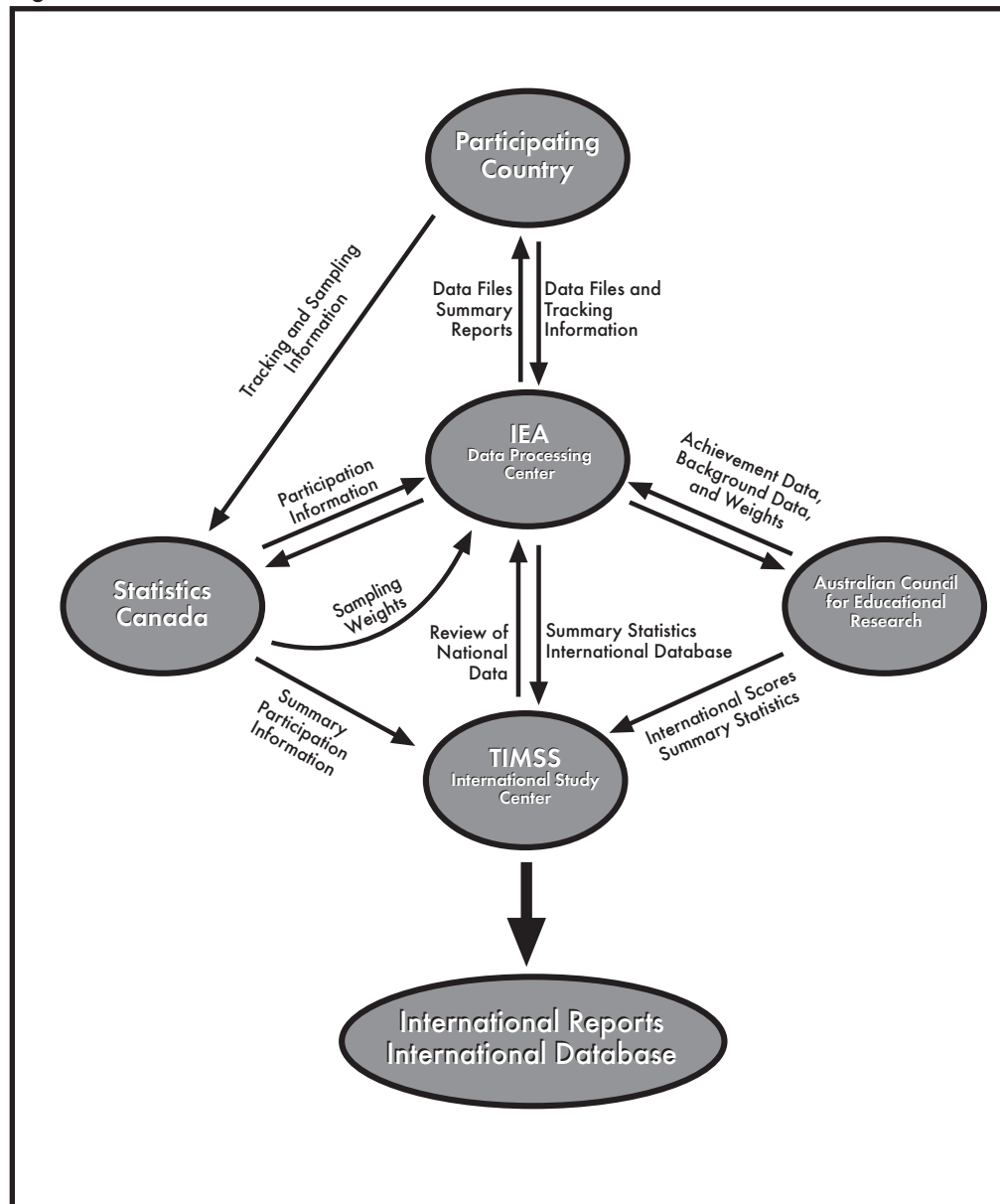
Once verified and in the international file format, the achievement data were sent to the Australian Council for Educational Research (ACER), where basic item statistics were produced for item review and an initial country-level scaling was conducted. An item review was undertaken by the staff at the TIMSS International Study Center (see Chapter 6). At the same time Statistics Canada received from the IEA Data Processing Center data files containing participation information for students in the sample. This information, together with information provided by the NRC, was used by Statistics Canada to calculate sampling weights, population coverage, and participation rates at the school and student level. The sampling weights were then sent to the TIMSS International Study Center for verification and forwarded to ACER to be used in the scaling. When the review of the item statistics was completed and the IEA Data Processing Center had updated the database accordingly, the revised data files were sent to ACER. ACER was then responsible for computing the international item difficulties and for scoring individual students on the international scales. Once the sampling weights and international scale scores were verified at the TIMSS International Study Center, they were sent to the IEA Data Processing Center for inclusion in the international database and distributed to the national research centers. The International Study Center prepared the international report tables and published the reports of the study results. A pictorial representation of the flow of the data files is presented in Figure 3.1.

A very important part of the data processing was the interaction among the staff at the TIMSS International Study Center, the staff at the IEA Data Processing Center, and the National Research Coordinators. At specific stages of the data verification, the IEA Data Processing Center returned countries' data files for checking. These data files were accompanied by computer printouts with summary statistics to be reviewed by the NRC, together with specific questions pertaining to the data.

3.2 DATA ENTRY AT THE NATIONAL RESEARCH CENTERS

Each TIMSS national research center was responsible for entering the achievement and background data into computer data files. Countries were provided with software adapted specifically for the purpose of TIMSS. The software, *DATAENTRYMANAGER* (DEM), was provided to each of the participating countries together with codebooks for data entry. The codebooks contained information about the variable names used for each variable in the survey instruments, and about field length, field location, labels, valid ranges, default values, and missing codes. The codebooks could be used together with DEM in the data entry process. Although this was the recommended

Figure 3.1 Flow of TIMSS Data Files



procedure, some of the participating countries elected to use a different data entry system. Data files were accepted from the countries provided they conformed to the parameters set in the international codebooks. In order to facilitate data entry, the codebooks and data files were structured to match the test instruments and questionnaires. This meant that there was a data file for each survey instrument.

Each country was responsible for submitting nine data files if participating fully in Population 1 (including performance assessment), and ten data files if participating fully in Population 2. Each of these files had its own codebook. The files for each population are listed in Table 3.1.¹ Although generally collected during the same session,

¹ "Written assessment" and "achievement" are used interchangeably to refer to the items or data from the test booklets administered to students. The file name for these data is "Written Assessment."

the student background data were entered separately from the student achievement data because the tests and questionnaires were administered as separate instruments. This was done to prevent students from looking back or ahead at their work in the achievement booklet and, most important, because the open-ended achievement items had to be scored following administration. Setting the system to enter the student background data in a file separate from the achievement data allowed the data manager from each country to start entering student background data without having to wait for the scoring process to finish.

Table 3.1 Files Submitted to the IEA Data Processing Center

File	Population 1	Population 2
Student Background	X	X
Written Assessment	X	X
Written Assessment Coding Reliability	X	X
School Background	X	X
Teacher Background	X	-
Mathematics Teacher Background	-	X
Science Teacher Background	-	X
Performance Assessment Student	X	X
Performance Assessment Coding Reliability	X	X
Performance Assessment School Tracking	X	X
Performance Assessment Student Tracking	X	X

The Student Background data file contains one record for each student in the sample, whether the student participated in the testing session or not. Entries were made in this file even if the student was excluded from the testing session. This file was used to record the information given by the students in the student questionnaire and other information on identification, participation, and sampling.

The Written Assessment data file contains one record for each student who was administered a test booklet. A record also was created for any student whose booklet was lost, but not for students who did not respond to the written assessment. The necessary information for these students was found in the Student Background data file.

In order to check the reliability of the free-response item coding, the free-response items in a random sample of 10 percent of booklets were coded independently by a second coder. The Written Assessment Coding Reliability file contains one record for each student whose responses to the free-response items were coded by a second coder.

The School Background data file contains one record for each originally sampled school, whether the school participated in the survey or not. They also contain records for those schools that participated in the survey as replacement schools. This file was used to register the information from the school questionnaire and on the participation status of schools.

The Teacher Background data file contains one record for each teacher listed as a teacher of a sampled student, even if the teacher was not administered a survey instrument. These files contain the information reported in the teacher questionnaires. The teachers for the third and fourth graders (Population 1) all received the same questionnaires with questions pertaining to the teaching of both mathematics and science. The teachers of the seventh and eighth graders (Population 2) received one of two questionnaires with questions regarding the teaching of either mathematics or science. The data for the mathematics teachers were recorded in a different file from the data for the science teachers.

The Performance Assessment Student data file created in each country contains one record for each performance assessment task that was assigned to a student, even if the student did not complete or attempt the task. Participating students are each represented with up to six entries in this data file depending on the number of tasks they were assigned to take.

The Performance Assessment School Tracking file contains one record for each school sampled for the performance assessment. This data file also contains the information recorded in the performance assessment Post Administration Form.

The Performance Assessment Student Tracking file contains one record for each entry in the Performance Assessment Tracking Form, so that each student is represented only once. This data file was meant to simplify the entering of the tracking information, which is of extreme importance for the linkage to the written assessment data. This file contains information about the specific tasks and task sequence assigned to each student.

The Performance Assessment Coding Reliability data file contains one record for each task that was coded by a second coder for reliability purposes.

Table 3.2 presents the total number of files and records of each type received from all the participating countries.

Table 3.2 Data Files Received by the IEA Data Processing Center

File	Population 1		Population 2	
	Files	Observations	Files	Observations
Written Assessment	27	206,662	43	338,908
Student Background	27	206,662	43	338,908
Written Assessment Coding Reliability	17	13,432	29	20,376
School Background	27	5,337	43	7,808
Teacher Background	27	10,757	-	-
Mathematics Teacher Background	-	-	41	13,885
Science Teacher Background	-	-	41	23,139
Performance Assessment	10	11,746	22	25,501
Performance Assessment Coding Reliability	4	641	14	1,852

In addition to submitting the data files, countries were also required to submit supporting documentation of their field procedures and copies of their national instruments (translated tests and questionnaires). The documentation included a report of their survey activities, a series of data management forms with clear indications of any changes made in the survey instruments or the structure of the database, and copies of all sampling tracking forms. Copies of these materials were archived at the IEA Data Processing Center and kept for reference purposes during data processing.

Each country was provided with a program called LINKCHK that was to be used to carry out checks on the data files prior to submitting them to the IEA Data Processing Center. The program was designed to help NRCs perform an initial check of the system of student, teacher, and school identification numbers after data entry, both within and between files.

LINKCHK performed checks for:

- Duplicate occurrences of identification numbers
- Inconsistencies in the identification numbering system
- Mismatches between different student files
- Mismatches between different teacher files
- Mismatches in the student-teacher linkage

Generally, two types of checks were made:

- Checks within the school, teacher, or student files
- Checks across linked files

The reports produced by the LINKCHK program allowed countries to correct problems in the identification system before transferring the data to the IEA Data Processing Center.

3.3 DATA CLEANING AT THE IEA DATA PROCESSING CENTER

Once the data were entered into data files at the national research center, the data files were submitted to the IEA Data Processing Center for checking and input into the international database. This process is generally referred to as data cleaning. The goals of the TIMSS data cleaning were to identify, document, and, where necessary and possible, correct deviations from the international file structure, and to correct key punch errors, systematic deviations from the international data formats, problems in linking observations between files, inconsistent tracking information between and within files, and inconsistencies within and across observations. The main objective of the process was to ensure that the data adhered to international formats and reflected accurately and consistently the information collected within each country.

Data cleaning involved several steps. Some of these were repeated in an iterative fashion until satisfactory results were achieved. During the first step of data cleaning, all incoming data files were checked and reformatted if necessary so that their file structure conformed to the international format. As a second step, all problems with identification variables, linkage across files, codes used for different groups of variables, and participation status were detected and corrected. The distribution for each variable was examined with particular attention to those variables that presented implausible or inconsistent distributions based on the information from the country involved.

During this stage, a series of data summary reports was generated for each country. The reports contained listings of codes used for each variable and pointed to outliers and changes in the structure of the data file. They also contained univariate statistics. The reports were sent to each participating country, and the NRC was asked to review the data and provide advice on how to best resolve inconsistencies in the data. In many cases the NRC was obliged to go back to the original booklets from which the data had been entered initially.

During the data cleaning process two main procedures were used to make necessary changes in the data. Errors due to incorrect data entry were usually corrected by keying the correct value directly. Inconsistencies in the hierarchical identification variables, whenever possible, were corrected by means of computer programs. In either case, all changes made in the data after they were received by the IEA Data Processing Center were documented. A database was created in which each change made in the data was recorded, and it was possible to reconstruct the original database received from a country.

In the following section each of the steps mentioned above is described in more detail.

3.3.1 Standardization of National File Structure

The first step in the data processing at the international level was to verify the compatibility of the national datasets with the international file structure as defined in the TIMSS international codebook. This was necessary before the standard cleaning with the Data Processing Center cleaning software could be performed.

Although the TIMSS international codebooks distributed with the data entry software gave clear and detailed instructions about the structure and format of the files each country was to submit to the IEA Data Processing Center, some countries opted to enter and submit their data files in other formats, using structures different from the international standard. For the most part, these differences were due to specific national circumstances.

The *TIMSS Guide to Checking, Coding, and Entering TIMSS Data* (TIMSS, 1995) asked countries to prepare and send their data files using the DEM software, which produces an extended dBase format. Some data files were also received in ASCII fixed format (raw data), SPSS format, and SAS format.

After the national files were converted into the extended dBase format, the structure of the files was inspected and deviations from the international file structure were identified. A standard software tool automatically scanned the file structure of the country files and reported the following deviations:

- International variables dropped
- National variables added
- Different variable length or number of decimal positions
- Different coding schemes or out of range values
- Specific national variables
- Gang-punched variables

Together with the inspection of the national data files, the data management and tracking forms submitted by each NRC were reviewed. As a result of this initial review, the Data Processing Center outlined and implemented necessary changes in the national data to make the files compatible with the international format. In most cases programs had to be prepared to fit the file structures and specificities of each country.

During this process some of the files were merged (for example, the Student Background and the Written Assessment data files). The structure of some of the files was also changed significantly, since direct correspondence to the instruments was no longer necessary. Some variables created during data entry for verification purposes only were not copied to the transformed data files. The changes made in the files during the cleaning process are described below. In general, variables used during data entry for verification were dropped from all files and new variables were added (e.g., reporting variables, derived variables, sampling weights, and achievement scores). What follows is a brief description of the changes performed in the files received from the countries.

3.3.1.1 Student Background File

Several new variables were added to the beginning of each record to represent students' participation status in the two testing sessions and in completing the student background questionnaire. The student's age computed from the date of testing and the date of birth were also added to the files, as were sampling weights and several achievement scores for both mathematics and science.

For Population 2, two versions of the student background questionnaire were available for administration. Each had its own data file and codebook. Although most countries chose to use one version of the questionnaire, some countries opted to use both versions. One version was tailored for educational systems where science is taught as an integrated subject (non-specialized version). The second version was tailored for edu-

cational systems where the sciences (biology, earth sciences, physics, chemistry) are taught separately (specialized version). Although a separate data file was created for data entry for each questionnaire version, these were then merged into one file with the same structure. This new file contained all variables from the version for non-specialized science teaching in the order in which they appear in the questionnaire, followed by all variables from the version for specialized science teaching that do not appear in the non-specialized version. For students who received the non-specialized version of the questionnaire, all questions that were given only in the specialized version were coded as “not administered.” For students who were assigned the specialized version of the questionnaire, all questions that were asked only in the non-specialized version were coded as “not administered.” The international structure of the Student Background data file is shown in Figure 3.2. In Population 1 there was only one version of the student questionnaire.

Figure 3.2 Revised Structure of the Student Background File

IDs	Tracking Information & Teacher Linkage	Background Information from the Non-Specialized Version	Background Information from the Specialized Version and not in the Non-Specialized Version	Weights	Scores	Derived Variables
-----	--	--	--	---------	--------	-------------------

3.3.1.2 Written Assessment File

The structure of the Written Assessment files prepared for data entry focused on the structure of the booklets (eight each for Populations 1 and 2). During data entry, once the version of the booklet was indicated, the data software displayed only the variables representing the items in that particular booklet. A variable was created for each item in a booklet, and the order of these variables reflected the order of the items within a booklet. This kept data entry and programming of the data entry software to a simple and rectangular structure. However, it also meant that a lot of redundant variables were created during data entry, since an item administered in more than one booklet was coded as a different variable for each booklet in which the item occurred. A useful feature of the redundancy is that it allowed the booklet administered to the student to be identified easily even if there was a key-punch error when the identification of the test booklet was entered.

After final cleaning, the Written Assessment files were restructured so that each item appeared in just one location in the student records, regardless of the test booklet it came from. This new structure reflects the item clusters used to assemble the booklets (Adams and Gonzalez, 1996) and not the booklet layout. The variables for the items that were not administered to the student were coded as “not administered.” The structure of the Written Assessment file is presented in Figure 3.3.

Figure 3.3 Revised Structure of the Written Assessment File

IDs	Tracking	Achievement Item Cluster					
		A	B	C	Y	Z

3.3.1.3 Written Assessment Coding Reliability File

The structure of the Written Assessment Coding Reliability file prepared for data entry also mirrored the structure of the eight test booklets. Again, a variable was created for each free-response item in a booklet, and the order reflected the order of appearance of the items within the booklets. In the final international data file the variables were re-arranged so that each item was represented by only one variable regardless of the booklet in which it appears. All other variables representing items not included in the booklet administered to the student were coded as “not administered.”

The final international version of the Coding Reliability file includes both the data from the 10 percent sample of students selected for reliability coding and the original data for these students. This enables the user of the Coding Reliability file to compare the codes without having to merge any files.

A third set of variables was included in the final international version of the file to reflect the agreement between the two codes assigned to the answers to the free-response items.

3.3.1.4 Teacher Background File

The structure of the Teacher Background files is similar to the that of the original data file used for data entry. For Population 2, two files were used for data entry, one corresponding to the mathematics teacher background questionnaire and one corresponding to the science teacher background questionnaire.

In some cases, a teacher taught more than one sampled class or course or, in the case of Population 2, both subjects to the same class or course. Although it would have been desirable to assign a questionnaire to a teacher for each class taught, in most countries the resulting burden to teachers was considered unacceptable. However, much of the information obtained from the questionnaires was not related to the specific class or course taught, but to background characteristics of the teacher (e.g., sex and age, teaching experience). This information was asked only once from the teachers.

Each teacher was assigned a unique identification number (Teacher ID) and a Teacher Link Number specific to each class taught by the teacher. The Teacher ID and Teacher Link Number combination identified a teacher teaching one specific class. For example, students linked to teachers identified by the same Teacher ID but different Teacher Link Numbers were taught by the same teacher but in different classes. If students were linked to a teacher observation identified by a combination of Teacher ID and

Teacher Link Number for which no data were obtained, but there was an observation in the teacher file with the same Teacher ID and a different Teacher Link Number with data available, all personal data for the teacher were transcribed to the missing observation. Thus, whether or not a teacher completed a questionnaire pertaining to a specific course, background information was sometimes available.

During data processing, teacher-related information was transcribed from other observations of the same teacher to teacher observations for which a questionnaire was not administered (or not returned). In some countries, more than two questionnaires per teacher were administered, but only one contained the personal information. In these cases, a similar transcription was made. Table 3.3 gives two examples of how teacher data have been transcribed.

Table 3.3 Examples of Teacher Data Transcribed to Files

Obs.	Teacher ID	Link No.	Class ID	Participation	Sections A & D	Sections B & C
					Teacher-related Data	Class-related Data
1	22201	1	22203	Questionnaire completed	Data	Data
2	22201	3	22205	No questionnaire assigned	Data transcribed from Obs. 1	No data
:	:	:	:	:	:	:
10	33302	4	33301	Questionnaire completed	Data	Data
11	33302	5	33302	Only class-related part completed	Data transcribed from Obs. 10	Data

3.3.1.5 School Background File

The file structure of the cleaned school data sets in the international database is identical to the structure used for data entry. No major changes were made. The file includes a School Identification number (ID) block and the variables in order of their appearance in the school questionnaire.

3.3.1.6 Performance Assessment Files

The structure of the Performance Assessment data files submitted by the national centers to the IEA Data Processing Center mirrored the structure of the instruments and tracking forms. To make the files suitable for further analysis, the Performance Assessment Student file was rearranged from a multi-record structure (i.e., multiple records for each student – one for each task taken by the student) to a single-record structure (i.e., one record per student). In addition, information from the Performance Assessment Student file and the Performance Assessment Tracking file were combined into one file, together with particular variables from the Student Background file (age, gender, achievement scores, etc.) and sampling weights computed by Statistics Canada.

This revised file was called the Performance Assessment Combined file. The Performance Assessment Coding Reliability files were kept separate and processed in the same manner as the Written Assessment Coding Reliability files

3.3.2 Standard Cleaning

After the data received from the countries were transformed into the international format, a set of standard cleaning rules were applied to each of the data files received from each country. These rules were applied using software the IEA Data Processing Center had developed to report and in many cases to correct inconsistencies in the data. Some inconsistencies could not be solved automatically but had to be reviewed carefully and appropriate corrections devised, where possible.

In particular, the following problems were sought and corrected whenever possible (for further details, please refer to Jungclaus and Bruneforth (1996)):

- Problems with identification, tracking, and other indicator variables
- Problems with split variables, i.e. variables where respondents were allowed to check more than one option
- Problems with the variable indicating the achievement booklet administered to the student
- Problems with filter and dependent questions

After as many problems as possible were solved at the IEA Data Processing Center (by reviewing the instruments and national documentation or by applying the cleaning rules), the Data Processing Center cleaning software was used a second time to create a report of remaining data problems. These reports were summarized and sent to the NRCs with specific questions and, in some cases, suggestions as to how the problem could be solved.

For the Performance Assessment files, the tracking data regarding the performance assessment rotation scheme, sequence number, and station participation status were compared with the tasks that students performed and for which data had been recorded in the Performance Assessment Student file. The tracking information also included the number of the written test booklet which each student had completed, thus enabling a double linkage check (in addition to the student ID) to the written assessment.

3.3.3 Item Cleaning

After applying the cleaning rules described above, the achievement data underwent a careful and detailed review.

For this purpose, an item analysis was performed using the item analysis software QUEST developed by ACER (Adams and Khoo, 1993). National scores in mathematics and science based on the Rasch model were calculated and several reports were generated with these data. Some data problems, such as items with changes in the coding scheme or switched response options, were detected and corrected at this point. Reports with summary item statistics were sent to the NRCs for their review.

The Coding Reliability data were compared with the Written Assessment data. For this purpose, the percentage of agreement between the codes assigned by the two coders was calculated on two levels: agreement between the number of score points assigned to an item and agreement on the two-digit diagnostic code.

After this initial review by the IEA Data Processing Center, reports were generated with item statistics. The TIMSS International Study Center used these reports to conduct a thorough review of the achievement item data. Details of this process are presented in Chapter 6 of this report.

3.3.4 Country-Specific Cleaning

Some of the anomalies detected by the checking procedure had to be solved case by case. During this process, it was important to find individual solutions that followed general guidelines, so that uniform solutions could be applied to similar problems in other countries.

The corrections made in this cleaning step were based on the NRCs' review of the preliminary statistics from the IEA Data Processing Center, the NRC field operations reports and instruments sent with the data, and the NRCs' comments on the data almanacs produced by the TIMSS International Study Center. In particular, the following steps were performed on a country-by-country basis to correct the data:

- Correcting switched options/categories in categorical background variables
- Deleting data entered for questions that were not included in the international versions of the questionnaires
- Deleting data entered in error
- Collapsing categories to match the international coding scheme
- Deleting data made incompatible by translation problems
- Copying data from one observation to another if the information requested was identical for both observations
- Adding dummy records to the files to ensure correct linkage across files

None of these steps were performed without the cooperation of the NRCs. They had to confirm or reject the suggested data changes; more important, in many cases they had to give detailed advice about the changes to be performed to the coding scheme.

3.3.5 Other General Cleaning

After transforming the data files into the international format, performing the standard cleaning on them, and reviewing the achievement data, two other kinds of checks were made: statistical checks and consistency checks.

3.3.5.1 Statistical Checks

Statistical checks were designed to find outliers for continuous variables, variables with very high percentages of missing values, and categorical variables with different numbers of options from the international version of the instruments. Statistical checks were performed separately for each country. For such checks, several preparatory steps were necessary. In particular, descriptive statistics were computed for each variable within each country and these statistics were stored in a database. The information compiled in this way was used as outlined below.

Outlier Detection

In order to check variables for extreme values, an outlier was defined as a value in a variable that is over 5 standard deviations above the mean for that variable, or with a value twice as large as the 90th percentile for the variable. Any such variables detected were carefully examined.

For some of the variables found by this procedure (e.g., number of students in a school), additional information was used to judge the plausibility of the detected outlying values. If the file contained obvious miskeys, the variable was coded to "Invalid" for the detected cases. Cases that could not be resolved at the Data Processing Center were reported to NRCs and treated according to their suggestions.

High Percentages of Missing Observations

Variables were flagged for investigation if more than 99 percent of the cases had missing values. If such a variable was detected, the corresponding question in the questionnaire was examined. Often in such cases the question was not completed by the respondents because it was not applicable. For example, teachers were asked a question about teaching the theory of relativity. Many teachers did not respond since relativity theory was not part of the curriculum in their country. Thus, the variables related to these questions show high missing rates. Another example would be that a question was not asked, but data entry errors gave the corresponding variable(s) inconsistent missing codes. In that case, the missing codes were made consistent.

Additional Response Options for Categorical Variables

The observed values for categorical variables were compared with the valid codes specified by the international codebook. If additional codes were found, the corresponding question in the questionnaire was examined. It was possible that the additional code was due to key-punch error during data entry. Where it was determined that this was the case, the corresponding categories were recoded to "Invalid." If, on the other hand, the question that was asked allowed additional categories, the NRCs were asked to help find a way to make the new code internationally comparable. If recoding was possible, the original value for the variable was kept in a separate country-specific variable. If it was not possible to recode to meet the international coding scheme, the original data were kept in a separate variable and the international variable was coded to an explicit missing code.

Response Options with a Frequency of Zero in Categorical Variables

If a frequency of zero was detected for an option of a categorical variable, the corresponding question in the questionnaire was checked as a precaution. If a category in the original version of the question was missing, the NRC was contacted to verify that the correct categories were retained. However, if the category was not missing in the questionnaire but was not checked by any respondent, the data were not changed. Quite often, variables belonging to groups of questions had zero frequencies for one or more of the categories. For example, the school questionnaire asked for the frequency of different types of student behavior in schools. Some forms of behavior did not happen often; thus the corresponding categories had a frequency of zero.

3.3.5.2 Consistency Checks

Consistency checks dealt with problems that were discovered in the first phase of the cleaning process, but not corrected at that time because information about the problems across countries was needed to decide on the rules to be applied. The following sections describe the checks applied to all countries and the inconsistencies that were corrected.

Student's Gender, Date of Birth, Age, and Date of Testing

If a student's sex as reported in the background questionnaire differed from that in the tracking information, the tracking version was replaced by the background questionnaire version in Population 2. In Population 1 the replacement was the other way around. The same substitution procedure was followed with regard to students' dates of birth. Changes in the date of birth were made provided that the value to be used in the substitution resulted in a valid age for the student. For students whose estimated age was less than ten years of age in Population 2, or less than six years of age in Population 1, the estimated age was coded as invalid.

If the date of testing was missing, it was replaced by the modal value of the student's class when available.

Teacher File

In the Teacher Files, two lists in the Population 2 questionnaires were considered and corrected separately: a list of subjects taught during a school week and a list of tasks that must be performed during a school week. If no zeroes were used, more than four variables in a list were coded differently from "Not administered," or values greater than zero could be found, then "Omit" codes were recoded to zero.

School File

The questions concerning the same course of instruction were checked for consistent answers. If all students followed the same course of instruction (filter = Yes) and the majority of answers was consistent with the filter, all answers in the "No" list were recoded to "Not applicable." If, on the other hand, valid answers could be found in the

“No” list and only missing values could be found in the “Yes” list, the filter was changed to “No.” Uncertain cases were reported and recoded directly if possible. Sometimes the appropriate response could be deduced from the answering pattern found in the data.

3.3.6 Performance Assessment Cleaning Routines

The Performance Assessment file cleaning routines were based on the data checks created for Written Assessment files, although some routines were modified to fit the structure of the Performance Assessment files. In addition, due to the design of the Performance Assessment and the linkages among the various files, it was necessary to develop special cleaning routines. These cleaning programs were of two types. One type of cleaning program flagged inconsistencies between the Performance Assessment Tracking file data and the Performance Assessment Student file data. The second type of cleaning program flagged problems associated with the Performance Assessment Combined file.

Performance Assessment cleaning problems could not be resolved automatically, but rather had to be solved case by case. It would have been very difficult to create general cleaning rules which could cover the complexity of the Performance Assessment design. The structure of Performance Assessment required case-by-case cleaning especially to resolve inconsistencies between the Performance Assessment Tracking file and Performance Assessment Student file. Problems were resolved by reviewing error report printouts and data, and through dialogue with the participating countries. All corrections were undertaken by editing the data files.

Similar to the written assessment items, the performance assessment item responses were analyzed with the QUEST program. Both Rasch statistics and classical item statistics were calculated, printed, and reviewed, as described in Section 3.3.4. The only difference to the procedure for the written assessment items is that all performance assessment item responses were scored using the two-digit coding scheme, like the open-ended items of Written Assessment.

The Performance Assessment Reliability Coding data were processed and statistics were produced for review in a similar manner as those for the written assessment.

After the Data Processing Center had reviewed all item statistics, they were sent to the participating countries and the International Study Center. Country-specific item statistics enabled NRCs to review their data, and international item statistics were sent to the International Study Center for an international review of all items for all countries.

3.4 DATA PRODUCTS

3.4.1 Data Almanacs

Together with their data files, each country received data almanacs produced by the TIMSS International Study Center that contained weighted summary statistics by grade, for each participating country, on each variable included in the survey instruments. There were two types of display. The display for *categorical variables* included

an estimate of the size of the student population, the sample size, the weighted percentage of students who were not administered the question, the percentage of students choosing each of the options on the question, and the percentage of students who did not choose any of the valid options. The percentage of students to whom the question did not apply was also presented in the almanac. For *continuous variables* the display included an estimate of the size of the student population, the sample size, the weighted percentage of students who were not administered the question, the percentage who did not respond, the percentage to whom the question did not apply, the mean, mode, minimum, maximum, and the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles. An example of such data displays is presented in Figures 3.4 and 3.5. These data almanacs were sent to each of the participating countries for review. When necessary, they were accompanied by specific questions about the data presented in them. These almanacs also were used by the TIMSS International Study Center during the data review and in the production of the reporting tables.

Figure 3.4 Example Data Almanac Display for Categorical Variable

```

1Third International Mathematics and Science Study
4:17 Sunday, September 21, 1997 1
Report on Student Background Variables - Population 2
Preliminary results: DO NOT CITE OR CIRCULATE

Question: Are you a boy or a girl? (BSBGSEX)
Location: SQ2-2

*****
1.SEVENTH GRADE
GEN\STUDENT'S SEX
1.GIRL 2.BOY Other
% % %
Country Population Sample %NA % % %
-----
Australia 238294 5599 1.3 52.0 48.0 1.3
Austria 89593 3013 3.3 53.2 46.8 3.6
Belgium (Fl) 64177 2768 0.3 49.4 50.6 0.3
Belgium (Fr) 49898 2292 1.3 53.2 46.8 1.8
Bulgaria 140979 1798 0.5 54.0 46.0 0.5
Canada 377732 8219 0.8 49.4 50.6 2.2
Colombia 619462 2655 0.9 49.9 50.1 1.1
Cyprus 10033 2929 0.2 48.9 51.1 0.3
Czech Republic 152492 3345 0.2 50.6 49.4 0.2
Denmark 44980 2073 5.1 51.2 48.8 5.1
England 465457 1803 1.8 45.7 54.3 1.8
France 860657 3016 3.4 49.6 50.4 3.5
Germany 742346 2893 0.8 50.9 49.1 1.5
Greece 130222 3931 0.2 48.2 51.8 0.4
Hong Kong 88591 3413 0.5 44.3 55.7 0.6
Hungary 118727 3066 2.0 50.4 49.6 2.4
Iceland 4212 1957 0.7 49.0 51.0 0.7
Iran, Islamic Rep. 1052795 3735 1.7 43.2 56.8 1.7
Ireland 68477 3127 1.1 54.0 46.0 1.1
Israel . . . . .
Japan 1562418 5130 0.0 48.4 51.6 .
Korea 798409 2907 0.2 42.4 57.6 0.2
Kuwait . . . . .
Latvia (LSS) 17041 2567 1.8 51.5 48.5 1.9
Lithuania 36551 2531 0.7 49.9 50.1 0.7
Netherlands 175419 2097 2.7 50.5 49.5 2.8
New Zealand 48508 3184 0.9 46.7 53.3 0.9
Norway 51165 2469 0.5 48.6 51.4 0.5
Portugal 146882 3362 0.6 51.6 48.4 0.6
Romania 295348 3746 0.5 51.8 48.2 0.5
Russian Federation 2168163 4138 0.2 51.0 49.0 0.2
Scotland 61938 2913 4.1 49.2 50.8 4.1
Singapore 36181 3641 0.4 49.9 50.1 0.4
Slovak Republic 83074 3600 0.0 50.9 49.1 0.0
Slovenia 28049 2898 0.2 51.3 48.7 0.2
South Africa 649180 5301 0.6 53.9 46.1 1.7
Spain 549032 3741 0.5 49.7 50.3 0.5
Sweden 96494 2831 0.5 48.8 51.2 0.6
Switzerland 66681 4085 0.6 49.7 50.3 0.6
Thailand 680225 5810 0.0 58.2 41.8 1.1
United States 3156847 3886 1.7 50.3 49.7 1.7

```

Figure 3.5 Example Data Almanac Display for Continuous Variable

14:17 Sunday, September 21, 1997 2

1Third International Mathematics and Science Study
 Report on Student Background Variables - Population 2
 Preliminary results: DO NOT CITE OR CIRCULATE

Question: Student's Age in Years (BSDAGE)
 Location: DERIVED

GRADE=1.SEVENTH GRADE

Country	Population	Cases	%Not Ad.	%Omit	%Not Ap.	Mean	Mode	Min	P5	P10	Q1	Median	Q3	P90	P95	Max
Australia	238294	5599	0.2	0.0	0.0	13.2	13.3	11.8	12.5	12.7	12.9	13.3	13.5	13.8	14.1	16.3
Austria	89593	3013	3.9	0.0	0.0	13.3	13.3	12.3	12.7	12.8	12.9	13.3	13.6	14.0	14.4	16.3
Belgium (Fl)	64177	2768	0.1	0.0	0.0	13.0	12.9	11.5	12.4	12.5	12.8	13.0	13.3	13.7	14.0	16.3
Belgium (Fr)	49898	2292	2.9	0.0	0.0	13.2	12.6	11.4	12.4	12.5	12.9	13.0	13.2	13.4	14.5	17.8
Bulgaria	140979	1798	0.5	0.0	0.0	13.1	13.2	11.2	12.6	12.7	12.9	13.1	13.3	13.4	14.5	18.0
Canada	377732	8219	0.2	0.0	0.0	13.1	13.1	10.2	12.4	12.5	12.8	13.0	13.3	13.8	14.3	18.0
Colombia	619462	2655	1.8	0.3	0.0	14.5	13.0	10.0	12.3	12.5	13.0	14.0	15.1	16.8	18.6	49.8
Cyprus	10033	2929	0.6	0.0	0.0	12.8	12.9	11.5	12.3	12.3	12.5	12.8	13.0	13.2	13.4	15.3
Czech Republic	152492	3345	0.0	0.0	0.0	13.4	13.3	10.8	12.8	12.8	13.1	13.3	13.7	13.9	14.3	16.3
Denmark	44980	2073	3.3	0.0	0.0	12.9	12.9	11.5	12.3	12.4	12.6	12.8	13.1	13.3	13.4	15.2
England	465457	1803	0.0	0.0	0.0	13.1	13.5	12.2	12.6	12.7	12.8	13.1	13.3	13.5	13.5	14.5
France	860657	3016	6.9	0.0	0.0	13.3	13.0	10.6	12.4	12.5	12.8	13.2	13.8	14.3	14.8	17.4
Germany	742346	2893	2.4	0.0	0.0	13.8	13.8	11.1	13.1	13.2	13.3	13.7	14.0	14.5	14.9	29.7
Greece	130222	3931	0.5	0.0	0.0	12.6	12.6	10.0	12.0	12.1	12.3	12.6	12.8	13.1	13.7	17.5
Hong Kong	88591	3413	0.5	0.2	0.0	13.2	13.3	10.3	12.5	12.5	12.8	13.0	13.3	14.0	14.6	18.6
Hungary	118727	3066	3.5	0.0	0.0	13.4	13.5	11.0	12.7	12.8	13.0	13.3	13.6	14.0	14.5	17.3
Iceland	4212	1957	0.1	0.0	0.0	12.6	12.8	10.4	12.2	12.3	12.4	12.7	12.8	13.0	13.1	15.1
Iran, Islamic Rep.	1052795	3735	10.9	0.0	0.0	13.6	12.7	11.2	12.7	12.7	12.9	13.4	14.1	15.0	15.7	18.3
Ireland	68477	3127	0.0	0.0	0.0	13.4	13.3	12.2	12.7	12.8	13.0	13.4	13.8	14.1	14.3	16.5
Japan	1562418	5130	0.5	0.0	0.0	13.4	13.4	12.2	12.9	13.0	13.2	13.4	13.7	13.8	13.8	15.7
Korea	798409	2907	0.0	0.0	0.0	13.2	12.8	10.5	12.8	12.8	12.9	13.2	13.5	13.7	13.7	14.9
Latvia (LSS)	17041	2567	0.2	0.1	0.0	13.3	13.0	12.0	12.7	12.8	12.9	13.2	13.6	14.0	14.4	16.4
Lithuania	36551	2531	0.0	0.0	0.0	13.4	13.1	12.0	12.8	12.8	13.0	13.3	13.6	14.0	14.4	16.2
Netherlands	175419	2097	0.4	0.0	0.0	13.2	12.9	10.3	12.6	12.7	12.8	13.2	13.5	13.9	14.3	16.5
New Zealand	48508	3184	0.0	0.0	0.0	13.0	13.1	11.2	12.4	12.5	12.8	13.0	13.3	13.5	13.6	15.1
Norway	51165	2469	0.0	0.0	0.0	12.9	13.0	11.2	12.3	12.4	12.6	12.8	13.1	13.3	13.3	14.8
Portugal	146882	3362	0.1	0.0	0.0	13.4	13.0	11.3	12.5	12.6	12.8	13.2	13.8	14.8	15.3	19.5
Romania	295348	3746	0.1	0.0	0.0	13.7	13.6	11.6	12.9	13.0	13.3	13.6	13.9	14.6	14.6	17.5
Russian Federation	2168163	4138	0.0	0.0	0.0	13.0	12.8	11.5	12.5	12.6	12.8	13.0	13.3	13.5	13.9	16.8
Scotland	61938	2913	0.5	0.0	0.0	12.7	13.0	11.3	12.2	12.3	12.4	12.7	13.0	13.1	13.2	14.3
Singapore	36181	3641	0.0	0.0	0.0	13.3	13.1	12.8	12.8	12.9	13.0	13.3	13.5	13.8	13.8	17.5
Slovak Republic	83074	3600	0.0	0.0	0.0	13.3	13.0	12.4	12.8	12.8	13.0	13.3	13.5	13.7	13.9	16.2
Slovenia	28049	2898	0.2	0.0	0.0	13.8	13.8	12.3	13.3	13.3	13.5	13.8	14.1	14.3	14.6	15.8
South Africa	649180	5301	5.4	0.0	0.0	13.9	13.0	10.0	12.0	12.1	13.0	13.4	14.9	16.1	17.3	24.8
Spain	549032	3741	0.0	0.0	0.0	13.2	13.0	12.4	12.4	12.5	12.8	13.1	13.4	14.3	14.8	16.3
Sweden	96494	2831	0.0	0.0	0.0	12.9	13.2	10.3	12.4	12.5	12.7	12.9	13.2	13.3	13.3	14.5
Switzerland	66681	4085	0.5	0.0	0.0	13.1	13.0	10.2	12.3	12.4	12.7	13.0	13.8	14.2	14.2	16.7
Thailand	680225	5810	1.8	0.0	0.0	13.5	13.6	11.6	12.5	12.8	13.2	13.5	13.8	14.1	14.6	15.1
United States	3156847	3886	0.0	0.0	0.0	13.2	13.1	10.2	12.5	12.7	12.8	13.2	13.5	13.9	14.3	16.2

3.4.2 Versions of the National Data Files

Building the international database was an iterative process. The IEA Data Processing Center provided NRCs with a new version of their countries' data files whenever a major step in data processing was completed. This also guaranteed that the NRCs had a chance to review their data and run their own checks to validate the data files.

Three versions of the data files were sent out to each of the countries before the TIMSS international database was made available. Each country received its own data only. The first version of the data files was sent to the NRC as soon as that country's data had been cleaned. These files contained national Rasch scores calculated by the Data Processing Center. Documentation, with a list of the cleaning checks and all corrections applied to the data, was included to enable the NRC to review the cleaning process. Univariate statistics for the background data and item statistics were also provided for statistical review of the data. A second version of the data files was sent to the NRCs when the weights and the international achievement scores were available and had been merged with the files. A third version of the data was sent together with the data almanacs after final updates had been made in the data files, to enable the NRCs to validate the results presented in the first international reports.

For the performance assessment, participating countries were provided with their performance assessment data as soon as they were cleaned and restructured. The data were distributed along with national item statistics and a codebook describing the new structure of the data.

When international weights and scores were available, each country received a new version of its performance assessment data and the International Study Center received the data for all countries.

3.4.3 Reports

Several reports were produced during data processing at the IEA Data Processing Center to inform and assist the NRCs, the TIMSS International Study Center, and other institutions involved in TIMSS. The NRCs were provided with diagnostic reports and univariate statistics to help them in checking their data. The TIMSS International Study Center and ACER were provided with international item statistics. The International Study Center also received international coding reliability statistics and international univariate statistics. A report was made to the TIMSS International Study Center and the TIMSS Technical Advisory Committee about each country's deviations and cleaning status as well as the major problems encountered during its data cleaning. The report also included general statistics about the number of observations per file and subpopulation and student response rates.

3.5 COMPUTER SOFTWARE

dBase was used as a standard database program for handling the incoming data. Tools for pre-cleaning and programs such as LINKCHCK (described earlier), and MAN-CORR and CLEAN (described below) were developed using CLIPPER for manipulating data and some data processing. Statistical analyses (e.g., univariate statistics) for

data cleaning and review were carried out with SAS. The final data sets were also created using SAS. For item statistics, the Data Processing Center used the QUEST software (Adams and Khoo, 1993).

The main programs that were developed by the Data Processing Center for TIMSS are described below. Most of the programs that were written for country-specific cleaning needs are not listed. Most of the programming resources in the main cleaning process were spent developing this set of programs.

3.5.1 MANCORR

The most time-consuming and error-prone part of data cleaning is the direct or “manual” editing of errors uncovered by the review process. Based on the Data Processing Center’s experience in the IEA Reading Literacy Study and the pilot phases of TIMSS, the data editing program MANCORR was developed. It is easy to use and generates automatic reports of all data manipulation. Its main advantage compared with other editors is that all changes in the data are documented in a log database, from which reports can be generated. As updated data were received from countries, the time-intensive manual changes could be automatically repeated. An “Undo” function allowed the restoration of original values that had been modified with the MANCORR program. The report on which changes were made in the data, by whom, and when was important for internal quality control and review. The MANCORR program was designed using CLIPPER in order to manipulate DATAENTRYMANAGER files.

3.5.2 CLEAN

The central program for data cleaning in TIMSS was the diagnostic program CLEAN, developed with CLIPPER. This program was based on the programs used in the IEA Reading Literacy Study and the TIMSS field tests. It checked all the TIMSS files separately, but also checked the linkages across files and made between-file comparisons. Then corrections were performed according to the rules described above (see Section 3.3.2 and, for a more detailed explanation, Jungclaus and Bruneforth, 1996). An important feature of the program is that it can be used on a data set as often as necessary. It could first be used to make automatic corrections, and subsequently for creating a report only, without performing corrections. Thus it was possible to run a check on the files at all stages of work until the file format was changed to the SAS format. This meant that the program was used not only for initial checks but also to check the work done at the Data Processing Center.

A feature of the TIMSS data cleaning tools is that all deviations are reported to a database, so that reports could be generated by type of problem or by record. Reports previously generated by the program could be compared automatically with newer reports to see which problems had been solved, and even more important, to see whether additional deviations were introduced during manual correction. Last, the databases (which included all reported deviations) were used to generate the final reports to be sent to the countries. These reports showed which deviations were initially in the data, which were solved automatically, which were solved manually, and which remained unchanged.

3.5.3 Programs Creating Meta Databases

Using SAS, several programs were developed by the Data Processing Center for reviewing and analyzing both the background data and the test items. For the background data, a meta database containing information provided by the initial analysis and by the international codebook was created. A meta database containing the relevant item parameters was also created for the achievement test items. Later, all statistical checks and reports used these meta databases instead of running the statistics over all data sets again and again. If the data for one country were changed, then statistics had to be recalculated only for this country; the tabulation program, which accessed only the meta database, could then be applied, since the other countries' values remained unchanged. This reduced the computing time for certain procedures from hours to a few minutes. Both databases are the base sources of several reports produced at both the national and international levels (e.g., for the univariate and item analysis reports).

The univariates and item statistics were prepared on a variable-by-country or country-by-variable basis to allow review at the national level and international comparison of individual variables.

3.5.4 Export programs

As mentioned above, SAS was the main program for analyzing the data. Using SAS, export programs were developed and tested to create output data sets for data distribution that are readable either by SAS or SPSS.

3.6 CONCLUSION

The structures and processes designed for the data processing of TIMSS, the largest international empirical educational study ever conducted, met the tremendous challenge provided. In planning for TIMSS data processing, the major problems were anticipated and provision for dealing with them incorporated into the data processing system. Even the most complicated school systems were handled adequately by the admittedly complex record identification system. This system had been criticized during the planning phase as too complicated, but it proved to be just complex enough to unambiguously identify observations and allow the linkage of files in every education system.

The Data Processing Center was closely involved in the planning phase of the study. This allowed the study to benefit from the Center's knowledge and experience in data processing. For example, it was anticipated that national adaptations and country-specific options would create problems not only during data processing but also in later analysis. Accordingly, international definitions were established that minimized such problems. Most of the problems encountered during data processing arose because countries sometimes modified the internationally-agreed procedures without notifying the Data Processing Center. The adaptation of record identification systems by some countries (because they felt the international system was too complex) created a lot of unexpected work.

Minor modifications, such as adding new categories to questions, switching the order of options, leaving out international response categories, or changing open-ended questions to multiple-choice questions, were easy to recode to match the international definitions unless countries completely restructured the questionnaires, resulting in the need for additional resources and energy to check and reorganize the data. This shows how important it is in any international study to verify translations of the national questionnaires and to ensure internationally comparable data.

Some problems arose due to communications difficulties. Early and continuous involvement of the data processing staff helped minimize the amount of time and work required, by the countries, the International Study Center, and the Data Processing Center, to produce clean data. It was very important that the data processing staff was easily accessible for the participating countries so that they could get help whenever they had problems. Modern technology, such as the capability to send facsimiles, as well as the Internet, makes the will to communicate, and not the distance between the participants, the most important factor in a successful study. TIMSS demonstrated this with the successful communication between the Data Processing Center in Hamburg, the TIMSS International Study Center at Boston College, Statistics Canada in Ottawa, and the Australian Council of Educational Research in Melbourne. The idea of a decentralized study proved feasible and workable. The time difference between the institutions involved occasionally even helped speed up the work: TIMSS was worked on around the clock.

REFERENCES

- Adams, R.J. and Gonzalez, E.J. (1996). The TIMSS test design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, Volume I: Design and development*. Chestnut Hill, MA: Boston College.
- Adams, R.J. and Khoo, S. (1993). *Quest: The interactive test analysis system*. Melbourne: Australian Council for Educational Research.
- Gonzalez, E.J. and Smith, T.A., Eds. (1997). *User Guide for the TIMSS international database: Primary and middle school years – 1995 assessment*. Chestnut Hill, MA: Boston College.
- Junglaus, H. and Bruneforth, M. (1996). Data consistency checking across countries. In M.O. Martin and I.V.S. Mullis (Eds.), *Third International Mathematics and Science Study: Quality assurance in data collection*. Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1995). *TIMSS guide to checking, coding, and entering the TIMSS data* (Doc. Ref.: ICC918/NRC449). Chestnut Hill, MA: Boston College.