

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

Eugenio J. Gonzalez  
*Boston College*

The procedures for processing the TIMSS data for Population 3 (final year of secondary school) were similar to those developed for Populations 1 and 2 (third and fourth grades, and seventh and eighth grades in most countries, respectively) and described in the *TIMSS Technical Report, Volume II* (Martin and Kelly, 1997). In accordance with the TIMSS reporting schedule, the Population 3 data were processed after the data for TIMSS Populations 1 and 2 were completed. The main procedural differences resulted from the absence of teacher questionnaires for the final-year population and that there were three achievement tests: (1) mathematics and science literacy; (2) advanced mathematics; and (3) physics. Otherwise, the data structure for Population 3 was identical to that of the younger populations.

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. While 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 Population 3 data is discussed in Chapter 7, the weighting procedures in Chapter 4, and the analysis and reporting in Chapter 8. This chapter describes the process followed in data entry, the creation of the international database, and the steps taken to ensure the quality and accuracy of the international database. It also describes the responsibilities of each participant in creating the international database. In particular, this chapter outlines the flow of the data files among the 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 data cleaning and provided to the national centers; and the computer software used in that process.

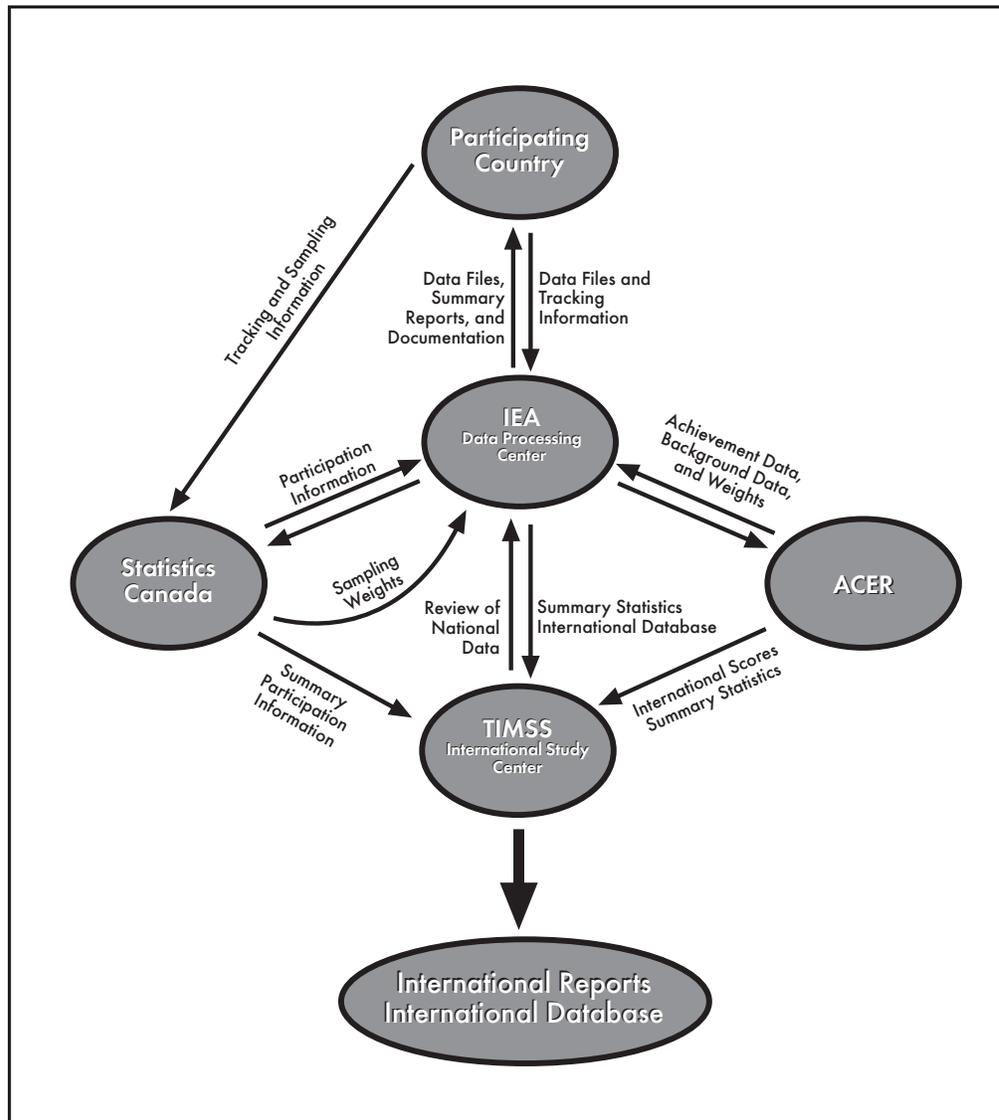
The TIMSS international database for Populations 1 and 2 was released for public use in September 1997, and that for Population 3 in June 1998. Both are available at the TIMSS website (<http://www.csteep.bc.edu./timss>) and through IEA Headquarters. The databases are each accompanied by a User's Guide and full documentation (see Gonzalez and Smith, 1997 and Gonzalez, Smith, and Sibberns, 1998).

### 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 main responsibilities of the IEA Data Processing Center at this point were to check that the data files submitted matched the international 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 finally, upon feedback from the individual countries and the TIMSS International Study Center, to construct the international database. The IEA Data Processing Center was also responsible 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. At this time, the staff at the TIMSS International Study Center undertook a thorough review of the item statistics (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. Figure 3.1 is a pictorial representation of the flow of the data files.

Figure 3.1 Flow of TIMSS Data Files



A very important part of the data processing was the interaction of 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 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 transforming the information from the achievement booklets and questionnaires into computer data files. Participating countries were provided with data-entry software adapted specifically for the purpose of TIMSS, DATAENTRYMANAGER (DEM), 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 were designed to be used with DEM in the data entry process. Although DEM was the recommended software, 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 four data files if participating in the testing of students in the final year of secondary school: Student Background, Achievement, Coding Reliability, and School Background. Each file had its own codebook.

Although generally collected during the same session, 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 of each country to start entering student background data without having to wait for scoring to finish.

The Student Background data file contains one record for each student in the sample. 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 Achievement data file contains one record for each student who was administered a test booklet. A record was also 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 contained 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 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 schools that participated as replacement schools. This file was used to register the information from the school questionnaire and on the participation status of schools.

Table 3.1 presents the total number of files and records of each type received from the participating countries.

**Table 3.1 Population 3 Data Files Received by the IEA Data Processing Center**

File	Files	Observations
Achievement	23	99107
Student Background	23	99107
Coding Reliability	14	9337
School Background	23	1282

In addition to 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 and tracking forms. 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 to carry out checks on the data files before 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. 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 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 advise how to best resolve inconsistencies. 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. Inconsistencies that could unambiguously be solved were corrected automatically by a data cleaning program. Errors that could not be solved using standard cleaning routines had to be solved case by case by the DPC staff. In either case, all changes made in the data were documented. A database was created in which each change 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, however, were 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.

As part of the file standardization process files were merged where applicable (for example, the Student Background and the Achievement data files) and the file structure was changed to facilitate data analysis, since direct correspondence to the instruments was no longer necessary. The changes made in the files during the cleaning process are summarized below. In general, variables created during data entry for verification were purposely 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 students' ages 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 advanced mathematics, physics, and mathematics and science literacy.

#### **3.3.1.2 Achievement File**

The structure of the Achievement files produced by each country reflected the structure of the nine test booklets. 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 it occurred. A useful feature of the redundancy is that it allowed the student's booklet to be identified easily even if there was a key-punch error when the identification of the booklet was entered.

After final cleaning, the Achievement 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) rather than 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 Achievement file is presented schematically in Figure 3.2.

**Figure 3.2 Revised Structure of the Student Achievement File**

IDs	Tracking	Achievement Item Cluster					
		A	B	C	...	K	L

**3.3.1.3 Coding Reliability File**

The structure of the Coding Reliability file prepared in each country also mirrored the structure of the nine 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 rearranged 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 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 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. The file includes a School Identification number (ID) block and the variables in order of their appearance in the school questionnaire. The major change was that school weighting variables and student achievement scores aggregated on the school level were added to the school file.

**3.3.2 Standard Cleaning**

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

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 for resolution.

### 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, physics, and literacy based on the Rasch model were calculated and several reports were generated with these data. Some data problems, such as items with inadvertent 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 achievement 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 the solutions could be uniformly 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 not internationally comparable due to 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, who 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 made in 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 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 than the international version of the instruments. Statistical checks were performed separately for each country. As preparatory steps, 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." 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. In some cases this was due to a data entry error.

### **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. The same substitution procedure was followed with regard to students' dates of birth. If the date of testing was missing, it was replaced by the modal value of the student's class when available.

**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.

**3.4 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, 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 3.3 and 3.4. 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 were also used by the TIMSS International Study Center during the data review and in the production of the reporting tables.

**Figure 3.3 Example Data Almanac Display for a Categorical Student Background Variable**

```

1Third International Mathematics and Science Study - 1995 Assessment      May 29, 1998  61
Student Background Variables - Students in the Final Year of Secondary School (INTMSL4=1)
Question: Do you intend to attend a four-year college or university? (CSBGFED4)
Location: SQ3-14A4
    
```

3. POPULATION 3							
GEN\CONT EDUC\ATTEND <4 YR COLLEGE							
U>							
Country	Population	Sample	%NA	1.YES %	2.NO %	NOT APP. %	Other %
Australia	170847	1941	2.0	65.7	25.5	8.8	2.9
Austria	70602	1962	3.7	32.1	41.1	26.7	4.4
Canada	263241	5232	0.7	62.5	33.7	3.7	5.3
Cyprus	4556	534	0.4	61.4	21.3	17.3	7.0
Czech Republic	137459	2167	0.4	24.6	35.8	39.6	0.8
Denmark	37872	2714	9.5	38.5	35.8	25.7	12.8
France	637935	1590	0.5	26.8	62.7	10.5	2.5
Germany	967705	2289	73.2	62.1	34.5	3.4	74.2
Hungary	111281	5091	3.2	9.3	74.0	16.6	12.0
Iceland	2308	1703	1.9	65.4	28.1	6.5	5.9
Israel	1357	1357	10.1	64.4	28.4	7.2	19.4
Italy	380834	1616	0.8	44.2	23.2	32.6	7.0
Lithuania	22161	2887	2.1	50.5	36.2	13.3	8.3
Netherlands	145916	1470	2.2	17.1	60.6	22.3	3.6
New Zealand	37549	1763	0.7	53.0	37.7	9.3	2.8
Norway	43806	2518	1.5	43.0	46.1	10.9	6.1
Russian Federation	1031187	2289	0.3	60.6	38.0	1.4	6.0
Slovenia	26644	1622	2.7	62.5	25.9	11.6	5.4
South Africa	374618	2757	2.1	75.2	19.1	5.7	8.8
Sweden	71243	3068	2.0	43.8	40.7	15.5	5.9
Switzerland	65140	3308	0.6	19.4	50.4	30.1	1.9
United States	2278258	5807	2.8	69.3	26.8	3.9	7.0

Figure 3.4 Example Data Almanac Display for a Numerical School Background Variable

May 29, 1998 20

Third International Mathematics and Science Study - 1995 Assessment  
 School Background Variables - Students in the Final Year of Secondary School (INTMSL4=1)

Question: How many classroom teachers are on the staff of your school? (FTE) (CCBGFTE4)  
 Location: SQ03-3D

\*POPULATION ID\*=3.POPULATION 3

Country	Population	Cases	%Not Ad.	%Omit	%Not Ap.	Mean	Mode	Min	P5	P10	Q1	Median	Q3	P90	P95	Max
Australia	170847	86	4.9	2.3	0.0	53.6	43.0	19.0	27.4	27.4	36.0	50.0	60.0	80.0	85.3	180.0
Austria	70602	169	8.5	3.0	0.0	45.0	45.0	4.0	13.3	23.0	35.0	59.4	86.0	110.0	120.0	385.0
Canada	263228	336	17.9	4.7	0.0	59.6	52.0	2.0	16.0	23.7	38.5	56.2	76.9	99.3	120.0	153.0
Cyprus	4556	28	12.9	0.0	0.0	52.6	29.0	20.0	20.0	29.0	41.7	51.4	61.0	75.0	76.0	113.0
Czech Republic	137459	150	0.2	1.5	0.0	29.3	50.0	1.0	9.0	10.5	19.0	28.1	37.3	49.0	53.0	91.0
Denmark	29095	91	0.8	1.8	0.0	60.4	46.0	6.5	28.0	36.0	52.0	62.0	73.0	80.0	85.0	110.0
France	637935	56	4.5	8.5	0.0	116.5	35.0	10.0	18.0	38.0	79.0	113.0	134.0	190.0	260.0	260.0
Germany	967705	152	35.3	3.7	0.0	56.5	48.0	2.0	15.0	25.0	34.0	53.0	75.0	103.0	106.0	112.0
Hungary	111281	204	15.0	5.8	0.5	47.4	35.0	1.0	15.5	21.0	34.3	46.0	60.5	76.5	82.5	94.0
Iceland	2308	30	5.7	10.2	0.0	43.1	8.5	8.0	8.5	10.0	33.2	44.6	55.0	59.0	77.5	77.5
Israel	1357	52	8.8	15.9	0.0	103.2	70.0	8.0	34.0	59.0	90.0	120.0	180.0	300.0	300.0	300.0
Italy	380834	101	1.4	8.9	0.0	78.3	69.0	9.0	15.0	30.0	48.0	73.0	107.0	130.0	154.0	190.0
Lithuania	22161	142	24.4	4.6	0.0	62.7	30.0	15.0	30.0	30.0	47.0	63.0	75.0	91.0	108.0	145.0
Netherlands	145916	79	100.0	0.0	0.0	*	*	*	*	*	*	*	*	*	*	*
New Zealand	37549	79	4.5	0.0	0.0	117.3	99.0	9.6	24.0	27.8	36.0	48.0	64.0	89.7	99.0	99.0
Norway	43806	131	9.1	3.1	0.0	58.2	40.0	16.0	23.0	28.0	40.0	51.8	70.0	97.0	120.0	160.0
Russian Federation	1031187	163	0.0	8.4	0.0	54.1	65.0	7.0	19.0	22.0	35.0	52.0	71.0	87.0	96.0	132.0
Slovenia	26644	79	66.9	2.6	0.0	58.2	33.0	18.0	31.0	32.0	41.0	55.0	74.0	89.0	91.0	152.0
South Africa	374618	90	69.9	0.7	0.0	23.3	16.0	1.0	1.0	12.0	16.0	23.0	28.0	40.0	51.0	51.0
Sweden	62244	126	4.5	13.4	0.0	84.6	70.0	3.0	13.0	37.0	61.0	80.0	110.0	134.0	148.0	190.0
Switzerland	65140	383	17.5	9.6	0.0	53.7	40.0	0.3	7.0	8.0	13.0	39.0	70.0	102.0	200.0	320.0
United States	2278258	211	17.1	5.7	0.0	71.2	42.0	1.0	13.0	19.0	35.5	63.0	93.0	126.0	156.0	263.0

### 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 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 for the achievement data were also provided for statistical review. 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, to enable the NRCs to validate the results presented in the first international reports.

### 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 the standard database program for handling the incoming data. Tools for pre-cleaning and programs such as LINKCHCK (described earlier), and MANCORR 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. The programming resources in the main cleaning process were spent largely in 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 made according to the rules described above. 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 making 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 can 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, 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. Another meta database containing the relevant item parameters was created for the achievement test items. Later, all statistical checks and reports used these 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 by either 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, successfully met the tremendous challenge. In planning for TIMSS data processing, the major problems were anticipated and provision for dealing with them was 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 barely general enough to identify observations unambiguously and allow the linkage of files in every education system.

The Data Processing Center was closely involved in the planning phase of the study. The study thus benefitted 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 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 to 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.
- Gonzalez, E.J., Smith, T.A., and Sibberns, H (Eds.). (1998). *User guide for the TIMSS international database: Final year of secondary school- 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.
- Martin, M.O. and Kelly, D.L. (Eds.) (1998). *TIMSS technical report, volume II: Implementation and analysis - primary and middle school years*. 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.