



Chapter 8

Creating and Checking the TIMSS 2003 Database

Juliane Barth, Ralph Carstens, and Oliver Neuschmidt

8.1 Overview

Creating the TIMSS 2003 database and ensuring its integrity was a complex endeavor requiring close coordination and cooperation among the staff at the IEA Data Processing Center (DPC), the TIMSS & PIRLS International Study Center (ISC) at Boston College, Statistics Canada, and the national research centers of participating countries. The overriding concerns were: to ensure that all information in the database conformed to the internationally defined data structure; that national adaptations to questionnaires were reflected appropriately in the codebooks and documentation; and that all variables used for international comparisons were indeed comparable across countries. Quality control measures were applied throughout the process to assure the quality and accuracy of the TIMSS data.

This chapter describes the data entry and verification tasks undertaken by the National Research Coordinators (NRC) and data entry managers of participating countries, the data checking and database creation procedures implemented by the IEA Data Processing Center in collaboration with the International Study Center and Statistics Canada, and the steps taken at all institutions to confirm the integrity of the international database. Section 8.2 describes the quality measures taken in order to document the comparability and consistency of the scoring of constructed-response achievement items within countries, across countries, and over time (from 1999 to 2003).

8.2 Creating and Checking the TIMSS 2003 Database

Database construction began with each national research center using the data-entry software and codebooks provided by the IEA DPC (see Chapter 6) to enter the data collected in the TIMSS 2003 survey into data files following the standard international format. Before sending the files to the DPC, national center staff applied a system of checks specified by the DPC to verify the structure and consistency of the data files. Checking and editing the national data sets was a matter of cooperation between the national centers, the ISC, Statistics Canada, and the DPC team.

On receipt of the data files from each country, the IEA DPC was responsible for checking their integrity, for applying standard cleaning rules to verify the accuracy and consistency of the data, and for documenting electronically any deviation from the international file structure. Any queries were addressed to the national centers and modifications were made to the data files as necessary. After all modifications had been applied, all data were processed and checked again. This process of editing the data, checking the reports, and implementing corrections was repeated as many times as necessary until all data were consistent and comparable within and between countries.

In preparation for creating the international database, the Data Processing Center provided item statistics to the national research centers while the International Study Center provided countries with data almanacs containing international univariate statistics so that National Research Coordinators could examine their data from an international perspective. This was one of the most important checks (in terms of international comparability of the data). While in a national context a particular statistic may seem plausible, it may become apparent in comparing data across countries that it is an outlier in an international context, even with accurate translation. Any such instances were addressed, and the corresponding variables either recoded or removed from the international database.

Once verified and in the international file format, the achievement data were sent to the International Study Center where basic item statistics were produced and reviewed. At the same time the Data Processing Center sent data files containing information on the participation of schools and students in each country's sample to Statistics Canada. This information, together with data provided by the National Research Coordinator from tracking forms and the *WinW3S: Within-School Sampling Software* (IEA, 2002a), was used by Statistics Canada to calculate sampling weights, population coverage, and school and student participation rates.¹

1 See Chapter 9 for details about TIMSS 2003 sampling design.

When the review of the item statistics was completed and the Data Processing Center had updated the database to include sampling weights, the student achievement files were sent to the International Study Center where the IRT scaling was conducted and proficiency scores in mathematics and science generated for each participating student.² Once the sampling weights and the proficiency scores were verified at the International Study Center, they were sent to the Data Processing Center for inclusion in the international database and then distributed to the national research centers.

8.3 Data Entry at the National Research Centers

Each TIMSS 2003 national research center was responsible for transcribing the information from the achievement booklets and questionnaires into computer data files. As described in Chapter 6, the IEA DPC supplied national centers with the Windows DataEntryManager (WinDEM) software and manual (IEA, 2002b) to assist with data entry and held a training session on the use of the software. The DPC also provided countries with codebooks describing the structure of the data. The codebooks contained information about the variable names used for each variable in the survey instruments, and about field lengths, field locations, labels, valid ranges, default values, and missing codes. In order to facilitate data entry, the codebooks and data files were structured to match the test instruments and international version of the questionnaires. This meant that for each survey instrument there was a corresponding codebook, which served as a template for creating the corresponding survey instrument data file.

To assist in applying the data-entry software to the TIMSS 2003 data, the International Study Center provided each national research center with a *Manual for Entering the TIMSS 2003 Data* (TIMSS, 2002a) detailing prescribed procedures for data entry and verification. In addition, the *TIMSS 2003 Survey Operation Manual* (TIMSS, 2002b) included general instructions about the test administration and the data entry process.

The data manager at the TIMSS national center in each country gathered data from tracking forms that were used to record information on students selected to participate in the study, as well as their schools, and teachers. Tracking form related information was entered with the help of the WinW3S sampling software distributed by the DPC (see Chapter 6). The responses from the student achievement booklets as well as student, teacher, and school questionnaires were entered into computer data files created from the codebook templates. While strongly encouraged to use the WinDEM software for data entry, a few participating countries elected to use a different data

² See Chapter 11 for details about scaling procedures.

entry system. However, they were required to conform to all specifications established in the international codebooks and to check their data with all consistency checks provided with the WinDEM software.

For each testing grade the following files were used during data entry:

- The WinW3S database contained sampling information as well as tracking form information (such as student's age, gender, and participation status) from all sampled students, teachers, and schools.
- The student background data file contained data from the *Student Background Questionnaire*. Additionally, these files contained tracking information for those countries, which did not use the WinW3S software.
- The student achievement data file contained the student's responses to whichever of the 12 test booklets was assigned to the student.
- In order to check the reliability of the constructed response item scoring, the constructed-response items were scored independently by a second scorer in a random sample of 100 of each test booklet type. The responses from these booklets were stored in a reliability scoring file.
- Because for eighth grade, separate *Mathematics Teacher* and *Science Teacher Questionnaires* were administered, two data files for the teachers' data were used, one for each questionnaire. For fourth grade, a single *Teacher Questionnaire* was administered, so data were entered into one teacher data file. For all countries not using WinW3S the data files also contained information from the teacher tracking forms.
- The school data file contained data from the *School Questionnaire*.

8.4 Data Checking and Editing at the National Centers

Before sending the data to the DPC for further data processing, countries were responsible for checking the data files with programs specifically prepared for TIMSS and for undertaking corrections as necessary. The first step was to apply the checking programs that are a feature of the WinDEM program. These checks are intended mainly to identify invalid data, but also can check the consistency between some basic variables. For example, an important feature of WinDEM is the ability to check that identification codes (IDs) are unique within a file. The WinDEM checks were mandatory for all countries. Additionally, after each file had been checked, the WinLINK program, which verifies the links between the various files, had to be applied. This software checks that the identification variables (student, teacher, class, and school identification codes) exist and match in related survey files. NRCs were required to resolve any problems identified by the within-country cleaning process before submitting data files to the IEA DPC.

8.5 Submitting Data Documentation to the IEA Data Processing Center

In addition to the data files described above, countries were requested to provide detailed data documentation to the IEA Data Processing Center. This included copies of all original survey tracking forms, copies of the national versions of test booklets and questionnaires, and a report of the survey activities. In order that all national adaptations to the survey instruments be documented, countries were required to submit *Data Management Forms and Cultural Adaptation Forms*.

Countries also were asked to send to the DPC the sample of test booklets selected for double-scoring the constructed-response items (around 1200 booklets altogether). The student responses to constructed-response items in these booklets will be digitally scanned and preserved for use in the next cycle of TIMSS in 2007, when they will be rescored by TIMSS 2007 scoring staff to monitor consistency in scoring practices between 2003 and 2007.

8.6 IEA DPC Quality Assurance Program

The IEA DPC went to great lengths to ensure that the data received from the TIMSS countries were of high quality and were internationally comparable. The foundation for quality assurance was laid before the first data arrived at the DPC through the provision to the TIMSS countries of software designed to standardize a range of operational and data-related tasks.

- The WinW3S software (IEA, 2002a) performed the within-school sampling operations adhering strictly to the sampling rules defined by Statistics Canada and the International Study Center. The software also created all necessary tracking forms and stored student- and teacher- specific tracking form information (such as student's age, gender, and participation status).
- The WinDEM program (IEA, 2002b) enabled key-entry of all TIMSS test and questionnaire data in a standard, internationally-defined format. The software also includes a range of checks for data verification.
- The WinLINK program (and LinkT03M, its DOS version) enabled NRCs to perform consistency checks on the identification variables across the TIMSS survey files.

A study as complex as TIMSS required a complex data cleaning design. To ensure that programs ran in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the following steps were undertaken:

- Before use with real data, all data-cleaning programs were thoroughly tested using simulated data sets containing all possible problems and inconsistencies.
- All incoming data and documents were registered in a specific database. The date of arrival was recorded, along with any specific issues meriting attention.
- The cleaning was organized following strict rules. Deviations in the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal.
- All corrections undertaken to country's data files were listed in a country-specific cleaning report.
- Occasionally it was necessary to make changes to a country's data files. Every such "manual" correction was logged using a specially-developed editing program (SAS-ManCorr), which recorded all changes and allowed DPC staff to undo changes, or to redo the whole manual cleaning process automatically at a later stage of the cleaning.
- Data Correction Software (DCS) was developed at the IEA DPC and distributed among the participating countries to assist them in identifying and correcting inconsistencies between variables in the background questionnaire files.
- Once the data-cleaning was completed for a country, all cleaning steps were repeated from the beginning to detect any problems that might have been inadvertently introduced during the cleaning process.
- All national adaptations that countries recorded in their documentation were verified against the structure of the national data files. All deviations from the international data structure that were detected were recorded in a "National Adaptation Database". This database is available for data analysts as an Appendix to the User's Guide to the TIMSS 2003 International Database.

8.7 Data Checking and Editing 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 main objective of the process was to ensure that the data adhered to international formats, that school, teacher, and student information could be linked between different survey files, and that the data accurately and consistently reflected the information collected within each country.

The program-based data cleaning consisted of the following steps:

- Documentation and structure check
- Identification variable (ID) cleaning
- Linkage check
- Resolving inconsistencies in background questionnaire data

8.7.1 Documentation and Structure Check

For each country, data cleaning began with an exploration of its data file structures and a review of its data documentation: Data Management Forms, Student Tracking Forms, Class Sampling Forms, Teacher Tracking Forms, and Test Administration Forms. Most countries sent all required documentation along with their data, which greatly facilitated the data checking. The DPC contacted those countries for which documentation was incomplete and obtained all forms necessary to complete the documentation.

The first checks implemented at the DPC looked for differences between the international file structure and the national file structures. Some adaptations (such as adding national variables, or omitting or modifying international variables) were made to the background questionnaires in some countries. The extent and nature of such changes differed across the countries: some countries administered the questionnaires without any changes (apart from the translations), whereas other countries inserted items or options within existing international variables or added entirely new national variables. To keep track of any adaptations, NRCs were asked to complete Data Management Forms as they adapted the codebooks. Where necessary, the DPC modified the structure of the country's data to ensure that the resulting data remained comparable between countries.

As part of this standardization process, since direct correspondence between the data-collection instruments and the files was no longer necessary, the file structure was rearranged from a booklet-oriented model designed to facilitate data entry to an item-oriented layout more suited to data analysis. Variables created purely for verification purposes during data entry were dropped at this time, and provision was added for new variables necessary for analysis and reporting (i.e., reporting variables, derived variables, sampling weights, and achievement scores).

After each data file matched the international standard as specified in the international codebooks, a series of standard cleaning rules were applied to the files. This was conducted using software developed at the IEA DPC that could identify and in many cases correct inconsistencies in the data. Each problem was recorded in a database, identified by a unique problem number

and with description of the problem and the action taken by the program or by the staff of the DPC.

Where problems could not be rectified automatically, they were reported to the responsible NRC so that original data-collection instruments and tracking forms could be checked to trace the source of the errors. Whenever possible, staff at the IEA Data Processing Center suggested a remedy, and asked the NRCs to either accept it or propose an alternative. Data files then were updated to reflect the solutions agreed on. Where the NRC could not solve problems by inspecting the instruments or forms, a general cleaning rule applying rectified these. After all automatic updates had been applied, remaining corrections to the data files were applied directly by keyboard, using a specially developed editing program (SAS-ManCorr).

8.7.2 Identification Variable (ID) Cleaning

Each record in a data file should have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. If two records share the same ID number, and contained exactly the same data, one of the records was deleted and the other remained in the database. If the records contained different data apart from the ID numbers, and it was impossible to identify which record contained the “true data,” both records were removed from the database. The DPC tried to keep such losses at a minimum, and in only a few cases were data actually deleted.

The ID cleaning focused on the student background questionnaire file, because most of the critical variables were present in this file. Apart from the unique student ID number, there were variables pertaining to the students’ participation and exclusion status – as well as dates of birth and dates of testing used to calculate age at the time of testing. The Student Tracking Forms³ were essential in resolving any anomalies, as was close cooperation with NRCs (in most cases, the Student Tracking Forms were completed in the country’s official language). The information about participation and exclusion was sent to Statistics Canada, where it was used to calculate students’ participation rates, exclusion rates, and student sampling weights.

8.7.3 Linkage Check

In TIMSS, data about students and their schools and teachers appear in several files. It was crucial that the records from these files be linked together correctly to obtain meaningful results. The linkage was implemented through a hierarchical ID numbering system incorporating a school, class, and student component,⁴ and is cross-checked against the tracking forms. The students’

³ Tracking Forms were used to record the sampling of schools, classes, teachers, and students. (see also chapter 6).

⁴ The ID number of a higher level is included in the ID number of a lower sampling level: the class ID includes the school ID, and the student ID includes the class ID (e.g., student 1220523 may be described as student 23 of class 05 in school 122).

entries in the achievement file and in the student background file must match one another; the reliability scoring file must represent a specific part of the achievement file; the teachers must be linked to the correct students; and the schools must be linked to the correct teachers and students.

8.7.4 Resolving inconsistencies in background questionnaire data

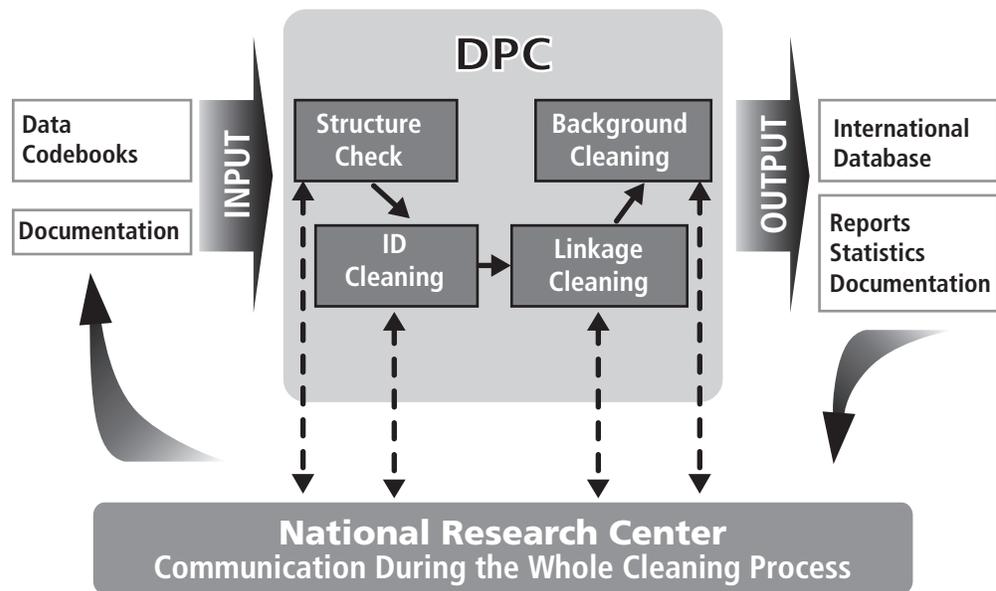
The number of inconsistent and implausible responses in background files varied from country to country, but no country's data were completely free of inconsistent responses. Treatment of these responses was determined on a question-by-question basis, using available documentation to make an informed decision. All background questionnaire data were checked for consistency among the responses given. For example, question number 2(a) in the School Questionnaire asked for the total school enrollment (number of students) in all grades, while 2(b) asked for the enrollment in the fourth grade only. Clearly, the number given for 2(b) should not exceed the number given for 2(a). All such inconsistencies that were detected were flagged and the NRCs asked to investigate. Those cases that could not be corrected or where the data made no sense were recoded to "Omitted".

Filter questions, which appear in some questionnaires, were used to direct the respondent to a particular section of the questionnaire. Filter questions and the dependent questions that follow were subject to the following cleaning rules: If the answer to the filter question was "No" or "Not applicable" and yet the dependent questions were answered, then the filter question was recoded to "Yes" or "Applicable."

Split variable checks were applied to questions where the answer was coded into several variables. For example question 5 in the Student Questionnaire listed a number of home possessions and asked the student to check all that applied. Student responses were captured in a series of 16 variables, each one coded as "Yes" if the corresponding possession was checked and "No" if left unchecked. Occasionally, students checked the "Yes" boxes but left the "No" boxes unchecked, or missing. Since in these cases it was clear that the unchecked boxes actually meant "No," these were recoded accordingly.

For further details about the standard cleaning procedures, please refer to the TIMSS *General Cleaning Documentation VII* (IEA, 2003).

Exhibit 8.1 Overview Data Processing at the DPC



8.7.5 National Cleaning Documentation

National Research Coordinators received a detailed report of all problems identified in their data, and of the steps applied to correct them. These included:

- Documentation of any data problems detected by the cleaning program and the steps applied to resolve them (*General Cleaning Documentation V11* (IEA, 2003))
- A record of all deviations from the international data-collection instruments and the international file structure

Additionally, the IEA DPC provided each NRC with revised data files incorporating all agreed edits, updates, and structural modifications. The revised files included a range of new variables that could be used for analytic purposes. For example, the student files included nationally standardized scores in mathematics and science that could be used in national analyses to be conducted before the international database became available.

8.7.6 Handling of Missing Data

When the TIMSS data were entered using WinDEM, two types of entries were possible: valid data values, and missing data values. Missing data can be assigned a value of omitted, not administered, or invalid during data entry.

At the IEA DPC, additional missing codes were applied to the data to be used for further analyses. In the international database, five missing codes are used:

- Not administered – the respondent was not administered the actual item. He or she had no chance to read and answer the question (assigned both during data entry and data processing).
- Omitted – the respondent had a chance to answer the question, but did not do so (assigned both during data entry and data processing).
- Logically not applicable – the respondent answered a preceding filter question in a way that made the following dependent questions not applicable to him or her (assigned during data processing only).
- Not reached (only used in the achievement files) – this code indicates those items not reached by the students, due to a lack of time (assigned during data processing only).
- Not interpretable (only used in the achievement files) – this code was used for multiple-choice items that were answered, but the chosen answer options were not clear, as well as for constructed-response items where the scorer assigned two or more scores (assigned during data entry and data processing).

8.8 Data Products

Data products sent by the IEA Data Processing Center to NRCs included both data almanacs and data files.

8.8.1 Data Almanacs and Item Statistics

Each country received a set of data almanacs, or summaries, produced by the TIMSS & PIRLS International Study Center. These contained weighted summary statistics for each participating country on each variable included in the survey instruments. The data almanacs were sent to the participating countries for review. When necessary, they were accompanied by specific questions about the data presented in them. They were also used by the International Study Center during the data review and in the production of the reporting exhibits.

Each country also received a set of preliminary national item statistics and reliability statistic reports for review purposes. The item statistics contained summary information about items characteristics, such as the classical item difficulty index, the classical item discrimination index, the Rasch item difficulty, and the Rasch mean square fit index. The reliability statistics contained summary statistics about the percent of agreement between scorers on the score assigned to the items.

8.8.2 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 country's 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. Before the TIMSS international database was published, three versions of the data files were sent to each country. Each country received its own data only. The first version was sent as soon as the data could be regarded as 'clean' concerning identification codes and linkage issues. These first files contained nationally standardized achievement scores calculated by the Data Processing Center using a Rasch-based scaling method. Documentation, with a list of the cleaning checks and corrections made in the data, was included to enable the NRC to review the cleaning process. 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 to the files. A third version was sent together with the data almanacs after all exhibits of the TIMSS International Report have been verified and final updates to the data files had been implemented, to enable the NRCs to validate the results presented in the first international reports.

8.8.3 The International Database

The international database incorporated all national data files. Data processing at the DPC ensured that:

- Information coded in each variable was internationally comparable
- National adaptations were reflected appropriately in all variables
- Questions that are not internationally comparable were removed from the database
- All entries in the database could be linked to the appropriate respondent – student, teacher, or principal.
- Sampling weights and student achievement scores were available for international comparisons

In a joint effort of the IEA DPC and the TIMSS & PIRLS International Study Center at Boston College, a National Adaptations Database containing all adaptations to questionnaires made by individual countries, and documenting how they were handled, was constructed. The meaning of country specific items can also be found in this database, as well as recoding requirements by the International Study Center. Information contained in this database was provided in the User Guide for the international database upon release of the TIMSS 2003 data.

The TIMSS 2003 international database is a unique resource for policy makers and analysts, containing student mathematics and science achievement and background data from representative samples of fourth and eighth grade students from 49 countries and four Benchmarking Participants.

References:

IEA (2002a). *WinW3S: Within-School Sampling Software*. Hamburg: IEA Data Processing Center.

IEA (2002b). *WinDEM: Software for data entry and verification*. Hamburg: IEA Data Processing Center.

IEA (2003): *General Cleaning Documentation V11*. Hamburg: IEA Data Processing Center

TIMSS (2002a), *Manual for Entering the TIMSS 2003 Data*, prepared by the International Study Center in cooperation with the IEA Data Processing Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 Survey Operations Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.