

CHAPTER 8

Creating the TIMSS International Databases

Kamil Kowolik
Milena Taneva

Introduction

The TIMSS International Databases are made publicly available and provide a valuable resource to those interested in secondary data analysis. The database published after each TIMSS assessment cycle includes scored responses for the assessment items and the context questionnaires administered during the cycle (e.g., Student, Home, Teacher, School, and Curriculum Questionnaire), plus variables for mathematics and science achievement scales and contextual scales. The International Database includes data for all countries and benchmarking entities participating in the cycle.

Preparing the International Database for each cycle and ensuring its integrity is a major undertaking, requiring extensive collaboration among IEA Hamburg, the TIMSS & PIRLS International Study Center, RTI International, the digital assessment platform providers, and the national centers of participating countries. In creating such an extensive international database, and given the diversity of participants, language, cultures, administration modes, and educational contexts from where the data originates, standardized and well-documented procedures must be followed in creating the International Database.

For each TIMSS assessment cycle, participating countries follow standardized survey operations procedures to prepare instruments, administer the achievement test, collect context questionnaire data, and prepare documentation to submit to IEA Hamburg for processing and validation. Data from digital instruments are captured or uploaded to servers for receipt by the IEA Hamburg data processing team, while data files for paper-based instruments must be prepared by the participating countries. Once the team at IEA Hamburg receives the TIMSS data and documentation, they follow standardized international data management and verification procedures to check for errors and inconsistencies and create standardized data and documentation outputs for each country. These procedures address the unique aspects for each instrument (e.g., achievement test and context questionnaires) and account for adaptations to the context questionnaires made for the particular contexts of the participating countries.

After the team at IEA Hamburg receives all data and documentation, data processing includes a uniform cleaning procedure for validation of the structure, identification variables, linkages, and context data. Before the international database is made available to the general public, three interim versions of the national database files are provided to each country for verification that their data are ready for analysis and publication.

The international data management and validation process implemented includes verification of the following:

- All information in the database conforms to a common international data structure.
- The content of all documentation accurately reflects national adaptations to questionnaires.
- All variables used for international comparisons are, in fact, comparable across countries (after harmonization, where necessary).

To maintain consistency across TIMSS assessment cycles in the interest of measuring trends, the same general procedures are applied in each cycle while updating some processes to improve efficiency and to accommodate changes to the assessment design, instruments, or other circumstances. In this context, IEA Hamburg is responsible for checking the data submitted by each country, applying standardized data cleaning rules to verify the accuracy and consistency of the data, and documenting any deviations from the international file structure. This process also includes importing and cleaning the data collected by the TIMSS Player that delivers the digital assessment to students, transferring student response data that require human scoring and student response data for automated scoring into their corresponding systems, and importing item scores for further processing.

National Research Coordinators (NRCs) from each participating country collaborate with IEA Hamburg to resolve any queries that emerge during the data cleaning process and check any interim versions of their database(s). NRCs also are provided with data almanacs that contain summary statistics for each variable. These can be used by the NRCs to validate their data. The TIMSS & PIRLS International Study Center conducts all operational psychometric analyses of the achievement and context questionnaire data, creating scale scores for the mathematics and science assessment, the context questionnaires, and other derived variables based on the questionnaire response data. The Sampling Team (RTI International, in collaboration with the Sampling Unit at IEA Hamburg) calculates the sampling weights.

The User Guide for the International Database, published with the data files for each TIMSS cycle, describes all data files and their variable contents, along with documentation about the achievement items and context questionnaire items. A supplement to the user guide provides the National Adaptations documentation for the TIMSS Context Questionnaires.

Preparing and Submitting National Data and Documentation

Data collected during the assessment comes from multiple sources and, therefore, requires coordination by participating countries before being submitted for processing and cleaning. There are three main sources of data: 1) the TIMSS Player, which delivers the digital assessment to students and collects student responses to the assessment; 2) IEA's Online SurveySystem (OSS) used to administer the context questionnaires; and 3) paper instruments in those countries that opted to administer the instruments on paper.

Data from the TIMSS Administration

The TIMSS assessment is designed to be delivered online through a web delivery system or offline on stand-alone PC devices using a USB delivery system. Whereas the online delivery system ensures that data can be immediately available on servers for further processing, the offline delivery system requires Test Administrators or School Coordinators to upload the student data after the testing session has concluded. National centers are provided with a data monitoring tool that lists all student records present on the upload server. This can be used to check and update data availability status based on the student sample selection.

Pre-Processing and Scoring Digital Data

Some pre-processing steps are required to prepare digital data in a suitable format for scoring and further processing. Data from the online or offline delivery uploads are received at IEA Hamburg in the form of JSON (JavaScript Object Notation) files, then transferred to the TIMSS & PIRLS International Study Center for automated scoring. The original JSON report structure contains multiple timestamped events for each student, recording the full history of “raw” student responses provided to each item (e.g., typed or numerical answers for constructed-response items or according to a predefined scheme for selected-response items).

JSON files are pre-processed and automated scoring is conducted at the TIMSS & PIRLS International Study Center (see [Chapter 7](#)), and the assigned item response codes are transferred to IEA Hamburg for data processing. At IEA Hamburg, the JSON files are converted to an SQL structure with a separate database for each country and grade level, including all data from the original file, identification variables related to the import of data, and additional fields for human-scoring purposes. Student responses to constructed-response items requiring scoring by humans are transferred to an online scoring system (IEA's CodingExpert) to be accessed by national teams of scorers. At the start of this process, it is essential that the scoring system does not contain any duplicate records. Once IEA Hamburg applies data processing procedures to merge any incomplete student records and has resolved any issues with duplicates, the scoring supervisors distribute responses to the scorers on their scoring teams. The scoring system is used by NRCs and their scoring staff to score the constructed-response items.

When scoring is completed, the scored response data for all achievement items are transferred to tables prepared for import into the data processing system.

Data from Online Questionnaire Administration

National centers administer the student, school, teacher, home, and curriculum questionnaires online through IEA's OSS. To ensure confidentiality and properly identify and validate respondents, national centers provide every respondent with individualized login information and information on how to access the online questionnaires.

Responses to questionnaires are collected online and stored directly on centrally-located servers. The online questionnaires may include skip logic, which minimizes response burden and improves data consistency. For example, a question asking "Does your school have a school library?" followed by "How often do you visit the school library?" will not ask the respondent the follow-up question if the respondent indicates there is no school library and will code the response to this question as "Not applicable."

Data Entry and Verification of Paper Instruments

In countries administering any instruments on paper, the national center is responsible for entering the responses collected with the paper instruments into data files using the IEA Data Management Expert (DME) software. DME is a software system developed by IEA Hamburg that facilitates manual data entry of responses to paper instruments and includes validation checks to identify inconsistencies. National centers are instructed to manually enter response data from any paper assessments or questionnaires that contain at least one valid response and to discard unused or empty instruments.

National centers enter responses from the paper instruments into data files using the DME and a predefined international codebook. The codebook defines the data structure to be entered and contains information about the names, lengths, labels, and missing codes of variables; valid response ranges for continuous measures or counts; and valid values for nominal or ordinal questions.

Participating countries are allowed to make national adaptations to certain questions in the international questionnaires (e.g., questions about parents' education are adapted to reflect national education systems). Countries making such adaptations are required to update the codebook structure to reflect the adaptations made to the national questionnaire versions before beginning the data entry process.

To ensure consistency across participating countries, the basic rule for manual data entry into DME requires national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning.

The guiding principles for manual data entry include the following:

- Responses to closed response items were coded as "1" if the first option was selected, "2" if the second option was selected, and so on.
- Responses to open-response questions, for example, the number of students in the sampled class, are to be entered "as is" even if the value is outside the originally expected range.

- Responses to filter questions and filter-dependent questions are to be entered exactly as filled in by the respondent, even if the information provided is logically inconsistent.
- Non-response, ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question) are coded as “omitted.”

As each respondent ID number is entered, the DME software checks it for alignment with a five-digit checksum generated by the Within-School Sampling Software (WinW3S). A mistype in the ID or the checksum results in an error message prompting the person entering the data to verify the entry. The data verification module of the DME also checks for a range of other issues, such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues are flagged by the software, the data entry staff are prompted to resolve the inconsistency before resuming data entry.

Paper Double-Data Entry

To assess data entry reliability, national centers are required to have a 5 percent sample of each survey instrument (e.g., achievement booklet or context questionnaire) entered a second time by a different data entry person, operating independently from the first. As part of the procedures, it is recommended that countries begin the double-data entry process as early as possible during the data capture period to identify possible systematic misunderstandings or mishandlings of data entry rules and to initiate appropriate remedial actions, such as retraining national center staff.

The acceptable level of disagreement between the originally entered and double-entered data is established at 1 percent or less for questionnaire data and at 0.1 percent or less for achievement data. Values above these levels require resolution of the discrepancy and re-entry of data. The level of disagreement between the originally entered and double-entered data is evaluated to ensure that the margin of error observed for processed data is well below the required thresholds.

Data Verification at the National Centers

Before confirming the completeness of the data submission and documentation for further processing, national centers carry out mandatory validation and verification steps on all collected data and undertake corrections as necessary. This is done by all countries, regardless of mode of administration.

The validation process is supported by an option in WinW3S to generate an inconsistency report. This report lists all types of discrepancies between variables recorded during the within-school sampling and test administration processes. It makes it possible to cross-check these data against the database for online respondents, the uploaded student data on the central international server, the online questionnaire database, and any paper-based data entered in the DME. While paper-based questionnaire data are being entered, the data manager or other staff at each national center can use the information from the tracking forms to verify

the completeness of the materials. Student participation information (e.g., whether a student participated in the assessment or was absent) is entered via WinW3S.

National data managers are requested to resolve such issues before confirming their final data submission. If inconsistencies remain or the national center cannot solve them, countries are asked to provide documentation on these problems.

Upon confirming submission of the validated data, including completion of any human scoring, NRCs also provide extensive documentation, including hard copies or electronic scans of all original Student and Teacher Tracking Forms, Student-Teacher Linkage Forms, and when applicable, a report on procedural activities collected as part of the online Survey Activities Questionnaire (see [Chapter 4](#)).

Data Processing and Data Cleaning

To ensure the integrity of the international databases, uniform and standardized data cleaning procedures are followed as part of processing national data, involving regular consultation with NRCs. The main objectives of the data cleaning process are to ensure that the data adhere to international formats, that school, teacher, and student information can be linked across different survey files, and that the data reflect the information collected within each country in an accurate and consistent manner.

After each country completes human scoring and confirms complete submission of data and documentation, a four-step cleaning procedure is conducted upon the submitted data and documentation:

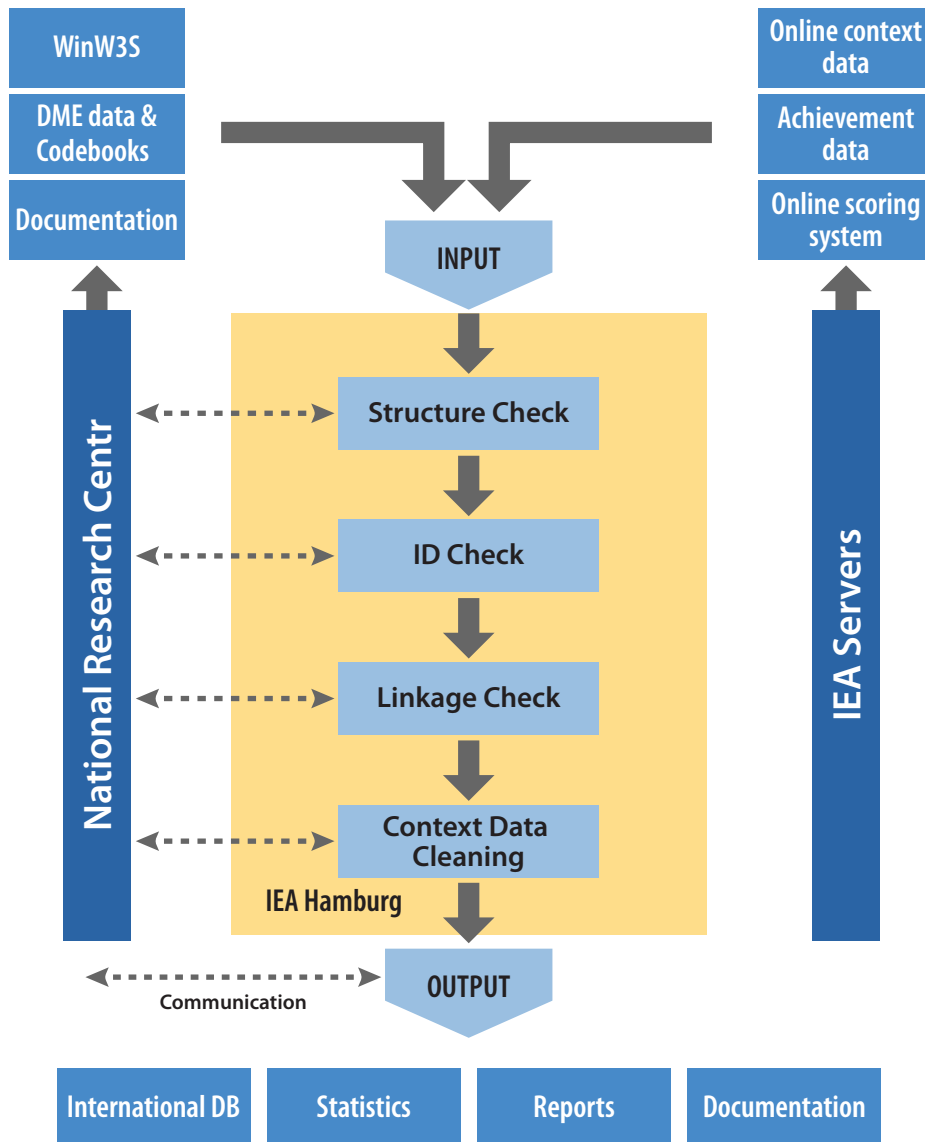
- checking documentation and paper-based data structure
- checking identification (ID) variables
- checking linkages
- context data cleaning

Data processing includes numerous iterations of the four-step cleaning procedure and is completed on each national data set in close collaboration with national centers. This iterative multi-step cleaning ensures that all data are properly cleaned and that any new errors that could have been introduced during the data cleaning are rectified. Any inconsistencies detected during the cleaning process are resolved in collaboration with national centers, and all corrections made during the cleaning process are documented in a cleaning report produced for each country. The cleaning process continues until all data are deemed consistent and comparable.

After the final cleaning iteration, sampling weights are calculated and added to the data files. Then, the data, including sampling weights, are readied for psychometric analysis. The NRCs are provided with interim data products to review at different points in the process.

As illustrated in Exhibit 8.1, the program-based data cleaning consists of a set of activities explained in the following subsections. IEA Hamburg carries out these activities in close communication with the national centers and the TIMSS & PIRLS International Study Center.

Exhibit 8.1: Overview of Data Processing at IEA Hamburg



Checking Structure and Documentation

For each country, data cleaning begins with a review of data file structures and its data documentation, including a review of National Adaptations Forms, Student Tracking Forms, Teacher Tracking Forms, Student-Teacher Linkage Forms, and the Survey Activities Questionnaire.

After the review, information from the tracking form is merged with the files, together with sampling information captured in the WinW3S database with the student-level database containing the corresponding student data from the achievement assessments. Data is merged from the student, school, home, and teacher questionnaires during this step. At this stage, data from the different sources are transformed and imported into one database to make this information available during all further data processing stages.

The first checks identify differences between the international and the national file structures. Some countries make adaptations (e.g., adding national variables or omitting or modifying international variables) to their questionnaires. The extent and nature of these changes differ across countries. Some countries administer the questionnaires without any modifications other than translations and necessary adaptations relating to cultural or language-specific terms, whereas other countries modify or add response categories within existing international variables or may add national variables.

National centers are asked to complete National Adaptations Forms to keep track of adaptations. In their adaptations, countries sometimes modify the structure and values of the international codebooks. In such cases, variables in the national data files are recoded to ensure that the resulting data remains comparable across countries. Details about country-specific adaptations to the international instruments can be found in a Supplement to the User Guide for the International Database for each TIMSS assessment cycle.

Once it is confirmed that each data file matches the international format, a series of standard data cleaning rules for further processing is applied. Processing during this step employs software checks developed by IEA Hamburg to identify and correct inconsistencies in the data. Each potential problem flagged at this stage is identified by a unique problem ID number and then described and recorded in a database. The action taken with respect to each problem is also recorded.

Problems that could not be rectified automatically are referred to the responsible NRC so that national center staff can check the original data collection instruments and tracking forms to trace the source of these errors. A remedy is suggested wherever possible, and national centers are asked to either accept it or propose an alternative. If a national center cannot solve the issue by verifying the instruments or forms, a general cleaning rule is applied to the files to rectify the error. When all automatic updates are applied, recoding scripts are used to apply any remaining corrections to the data files directly.

Checking Identification Variables

Each record in a data file needs to have a unique identification number. Except in the case of multiple student logins in a digital achievement test session (e.g., in case of a device or system crash), the existence of records with duplicate ID numbers in a file implies an error of some kind. When two records in a database share the same ID number and contain exactly the same data, one of the records is deleted, and the other one is kept in the database. In the rare cases where records with the same ID contain different data and it is not possible to identify which record contained the more reliable or complete version of the data, national centers are asked which record to keep.

Although the ID cleaning covers all data from all instruments, it focuses mainly on the student data files. In addition to checking the unique student ID number, it is crucial to check variables pertaining to student participation and exclusion status. Confirming students' birth dates and dates of testing is also crucial in order to correctly calculate student age at the time

of testing. The Student Tracking Forms provide an important tool for resolving anomalies in the database.

Checking Cross-File Linkages

As data on students, parents, teachers, and schools appear in a number of different data files, a process of linkage cleaning and verification is implemented to ensure that the data files can be correctly linked together. The linking of the data files follows a hierarchical system of identification codes that includes school, class, and student components. These codes link the students with their class and school membership.

Linkage cleaning consists of a number of checks to verify that student entries match across achievement files, student context questionnaire data files, scoring reliability files, and home background files. In addition, at this stage, checks are conducted to ensure that teacher and student records link correctly to the appropriate schools. The Student Tracking Forms, Teacher Tracking Forms, and Student-Teacher Linkage Forms are crucial in resolving any anomalies. During this process NRCs are consulted about any problematic cases, providing the national centers with standardized reports listing all inconsistencies identified within the data.

All cleaning procedures are always conducted in close cooperation with the national centers. After the national center staff verifies the accuracy of the identification variables and linkages, student sampling weights are calculated.

Context Data Cleaning: Resolving Inconsistencies in Context Questionnaire Data

The amount of inconsistent and implausible responses in questionnaire data files can vary considerably across countries. The treatment of inconsistent responses is determined question-by-question, using all available documentation to make informed decisions. All questionnaire data is checked for consistency across the responses given. For example, the response to a question asking for enrollment in the target grade should not exceed the response to the question asking about total school enrollment. Similarly, the response to a question asking a teacher how many years they spent in the school as a teacher should not exceed the years they have been teaching. Inconsistencies of this kind are flagged, and the national centers are asked to review these issues. Cases that cannot be corrected are coded as “invalid.”

Filter questions, which appear in some questionnaires, direct respondents to a particular set of questions only applicable to a subset of respondents. In online questionnaires, the system automatically redirects the respondent to applicable questions. In paper-based questionnaires, the respondent is given instructions to skip items, but it is possible they provide answers anyway. In these cases, the following cleaning rule is applied to these filter questions and their dependent questions. If a respondent answered “No” to a filter question such as “Does your school have a school library?” but responded to the dependent question “About how many books are there in the school library?”, the dependent question is recoded as “logically not applicable.” Using the same questions as an example, if the response to the filter question was

omitted but a response was provided to the dependent question, the filter question is recoded to “Yes.”

During this process, split variable checks are also applied to questions where the answers are coded into several variables. For example, when a student is asked “Do you have any of these things at your home?”, there are several choices provided, with the expectation that the student will indicate “Yes” or “No” for each of the options. If only “Yes” has been selected for some of them, all others are considered to be “No” answers.

In addition, student reports to items on age in the student questionnaire are checked against the tracking information provided by the School Coordinator or Test Administrator during the within-school sampling and test/questionnaire administration process. When information on birth year and month is missing in the student questionnaire, this information is copied over from the tracking data to the questionnaire. If discrepancies are found in age between student questionnaire responses and existing tracking data, the information is verified with the national center, and the national center investigates which source of information was correct. If unresolved, tracking data is used rather than questionnaire data at the fourth grade, and questionnaire data is used rather than tracking data at the eighth grade.

Handling of Missing Data

Two overarching types of entries are possible during the TIMSS data capture: valid data values and missing data values. IEA Hamburg applies additional missing data codes to facilitate further analyses. This process leads to four distinct types of missing data in the international database:

- **Omitted or invalid:** The respondent had a chance to answer the question but did not do so, leaving the corresponding item or question blank. This code is also used if the response was uninterpretable or out-of-range.
- **Not administered:** The item or question was not administered to the respondent, which meant that the respondent could not read and answer the question. The “not administered” missing code is used for those student test items that were not in the set of assessment blocks administered to a student, known as missing by design (due to the rotation of assessment blocks), or in rare cases, due to technical failure or incorrect translations. This missing code is also used for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires. For example, this situation applies to home questionnaire data for students who participated in the student test but whose parent/guardian did not answer the home questionnaire. In addition, the not administered code is used for individual questionnaire items that a national center decided not to include in the country-specific version of the questionnaire.
- **Logically not applicable:** The respondent answered a preceding filter question in a way that made the following dependent questions not relevant.

- **Not reached:** This is used only for individual items of the student assessment. This code applies to those items that the student did not attempt at the end of the booklet, either because time ran out or because the student stopped responding. “Not reached” codes are derived as follows: First, the last sequential response given by a student in a session is identified. Then, the first response after this last answer is coded as “omitted.” Finally, all following responses to these items in the session are coded as “not reached.”

Data Cleaning Quality Control

TIMSS is a large and highly complex study with the highest standards for data quality. Maintaining these standards requires extensive interrelated data-checking and data-cleaning procedures. To ensure that all procedures are conducted in the correct sequence, that no special requirements are overlooked, and that the cleaning process is implemented independently of the persons in charge, the data quality control process includes the following:

- **Thoroughly testing all data cleaning programs:** Before applying the programs to real data, they are used on simulated datasets containing all possible problems and inconsistencies.
- **Registering all incoming data and documents in a dedicated database:** The date of arrival and specific issues requiring attention are logged. For digital data, the delivery applications automatically track dates for responses and submissions. Paper-based data must be manually logged.
- **Carrying out data cleaning according to strict rules:** Deviations from the cleaning sequence are not possible, and the scope for involuntary changes to the cleaning procedures is minimal.
- **Documenting all systematic data recoding applied to all countries:** All changes to data in the comprehensive cleaning documentation provided to national centers are recorded.
- **Logging every “manual” correction to a country’s data files in a recoding script:** Logging these corrections, which occur only occasionally, allows for undoing specific changes or redoing the whole manual cleaning process at any later stage of the data cleaning process.
- **Repeating, on completion of data cleaning for a country, all cleaning steps from the beginning:** This step allows for detecting any problems that might have been inadvertently introduced during the data cleaning process.
- **Working closely with national centers at various steps of the cleaning process:** National centers are provided with the processed data files and accompanying documentation so that center staff can thoroughly review and correct any identified inconsistencies.

National adaptations recorded in the documentation for the national datasets are compared with the structure of the submitted national data files. Any identified deviations from the international data structure are recorded in the National Adaptations Database and for the supplementary materials provided with the User Guide for the International Database for each TIMSS assessment cycle. Whenever possible, national deviations are recoded to ensure consistency with the international data structure.

Interim Data Products

Before the International Database is finalized, three major interim versions of the data files are sent to each country, with additional interim versions as necessary for the cycle. The dates when these versions are provided depend on when countries administer the assessment. Documentation, with a list of the cleaning checks and corrections made in the data, is included with the data files. For the three major interim versions, each country only receives its own dataset and does not have access to other countries' data.

The first version of the interim data files is sent to the countries as soon as the data can be considered “clean” regarding identification codes, linkage issues, and context data inconsistencies. A second version is sent to the countries when all national adaptations and the feedback resulting from the review of the first version are implemented. NRCs are asked to confirm that the data are ready for the operational psychometric analysis. A third version of the data files is sent to the countries when the weights and achievement scale variables are available and have been merged with the data files. This version is structurally equivalent to the files to be published in the final TIMSS International Database. It contains only those student records used in the final analysis and satisfy the sampling standards.

Interim data products are accompanied by detailed data processing and national adaptation documentation, codebooks, and summary statistics. The summary statistics include country-by-country univariate statistics for all context questionnaire variables. For categorical variables, the percentages of respondents choosing each of the response options are displayed. For continuous variables, various descriptive statistics are reported, including the minimum, maximum, mean, median, mode, and percentiles. For both types of variables, the percentages of missing data are reported.

Additionally, item analysis and reliability statistics are provided for the achievement items, including information such as the number of valid cases, percentages, percentage correct, Rasch item difficulty, and scoring reliability. These statistics are used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, etc. More information on item almanacs and reviewing item statistics is available in [Chapter 10](#).

Final Product—the TIMSS International Database

The extensive data cleaning effort described in this chapter helps ensure that the TIMSS International Database for each cycle contains high-quality, internationally comparable data. More specifically, the process ensures that:

- Information coded in each variable conforms to the international scheme.
- National adaptations are reflected appropriately in all variables.
- All entries in the database can be successfully linked across students, teachers, and schools.
- Sampling weights and student achievement variables are available for international comparisons.