Chapter 9

Post-collection Data Capture, Scoring, and Processing

Alena Becker

Abstract The International Association for the Evaluation of Educational Achievement (IEA) collects large-scale assessment data from multiple countries and in many languages. Three important processes are data capture, scoring and coding, and data processing, which are supported by rigorous quality control mechanisms that IEA uses to ensure data comparability across countries. This chapter describes the steps that IEA takes to ensure the reliability of the data. Depending on the scope of the study, IEA uses different data capture systems during or after data collection. While data from multiple choice items can be captured directly, other data (e.g., responses from open ended questions) must be coded or scored uniformly to ensure comparability across countries and safeguard quality. Data processing starts only after data capture, scoring, and coding of data are finalized. As part of its data quality measures, IEA provides training for national representatives, instructional manuals, and study-specific software products to participating countries. Check protocols included in the software facilitate adherence to technical standards and minimize errors. Some data, while collected according to national conventions, ultimately needs to conform to IEA international formats. Moreover, the data must be assembled in a way that enables three linkages: schools with countries, teachers and classes with schools, and students with their teachers and parents. Data processing also detects and recodes any deviations from international formats or insufficient links between respondent levels.

Keyword Coding · Data processing · International database · Manual data capture · Online data collection · Paper data collection · Quality control · Reliability · Scoring

A. Becker (✉)

International Association for the Evaluation of Educational Achievement (IEA), Hamburg, Germany

e-mail: alena.becker@iea-hamburg.de

9.1 Introduction

Reliability is critical when judging the adequacy and quality of an assessment. International assessments, because of their complexity and design, place additional demands on developers and test administrators to ensure that the concerns related to the reliability of the data are addressed. The assessments developed by IEA share many common features; the more complex assessments, such as the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS), require response data to be captured from school principals, teachers, students, and sometimes parents or guardians. These data need to be captured systematically and accurately across multiple countries, multiple languages, and often, multiple populations within countries. Some data, while collected according to national conventions, ultimately need to conform to the international formats and have to be assembled in a way that links schools to countries, teachers and classes to schools, and students with their teachers and parents. Moreover, each student receives one specific cognitive instrument out of a pool of multiple instruments. To reduce the length of the cognitive instruments, items are arranged in blocks, which are rotated through multiple different instruments. To achieve a balanced number of responses across all items the cognitive instruments are evenly assigned to students by the IEA Within School Sampling Software. As IEA transitions from paper-and-pencil testing to computer-based assessment (CBA), data capture, processing, and scoring are evolving to reflect the new modalities. However, because the readiness of countries to adopt CBA differs, paper-and-pencil testing and CBA often have to operate in parallel.

To meet the challenge of ensuring high quality and comparable data, uniform methods of data capture and scoring have to be applied, first by country representatives at the national centers and later during the international data processing undertaken by IEA (Fig. 9.1). To facilitate these processes, IEA has developed training and standardized data capture and scoring procedures, together with specialized software; these include the IEA Data Management Expert (IEA DME), the IEA Windows Within School Sampling Software (IEA WinW3S), IEA Coding Expert, IEA eAssessment System, IEA Online SurveySystem, and the data processing programs used by IEA to process the data (Table 9.1). These procedures and software products are continuously reviewed to ensure the best quality possible.

An essential component of IEA training for participating national research coordinators (NRCs) involves an initial field trial that follows the same procedures as the subsequent main data collection. The field trial is a smaller version of the main data collection, collecting data from a small sample of respondents. It helps to identify any potential weaknesses in the procedures and rules, enabling these to be managed prior to the main data collection and thus further improving the reliability of the data management processes.

The transition from paper-and-pencil to online data capture began in 2004, with data from context questionnaires collected online via the IEA Online SurveySystem, software specifically designed for this purpose. As more countries expressed interest

Fig. 9.1 Process overview of data capture, scoring, and processing

Table 9.1 Overview of IEA software used during and after data collection

Software	Purpose
IEA Windows Within School Sampling Software (IEA WinW3S)	Used to perform the within-school sampling of classes, teachers, and students according to study specifications. Tracks the participation status of respondents and monitors the response rates during administration
IEA eAssessment System	Used to create and translate context questionnaires and test booklets and to collect data online. Can be used to create paper instruments for paper administration
IEA Online SurveySystem	Used to create and translate context questionnaires and to collect data online
IEA Data Management Expert (IEA DME)	Used to manually enter data from paper instruments
IEA Coding Expert	Used to code and score open-ended questions
Windows Data Entry Manager (WinDEM)	Used to manually enter data from paper instruments

in transitioning from paper-and-pencil testing to CBA, IEA developed additional software to enable this (the IEA eAssessment System). The relatively new and exciting field of computer- and tablet-based testing has allowed item developers to construct more interactive items and to assess new sets of competencies.

However, the transition from paper-and-pencil testing to CBA and the expansion of item types present challenges for quality assurance, reliability, and comparability over time. Developers have to assess whether the different modes of data collection produce different results. For example, Fishbein et al.’s (2018) integrated mode effect

study for TIMSS 2019 demonstrated that the different modes of data collection in paperTIMSS and eTIMSS did not produce strictly equivalent results; they proposed modifications to the item calibration model used for TIMSS 2019 so that TIMSS trend measurements could be maintained.

Quality control occurs both at the national center and at the IEA. Some of the quality control procedures and checks that are implemented at the national center are repeated at the IEA using the same software and the same procedures. This ensures that the national quality control has been implemented as requested. Additional quality control procedures and checks with different software take place at IEA.

9.2 Manual Post-collection Data Capture and Management Training

For those assessments that rely on paper-and-pencil administration, data collected from respondents needs to be captured and transformed into an electronic format. All participating countries receive training from IEA on how to execute the data entry procedures at international data management training sessions. These training sessions are conducted twice per study cycle, once before the field trial and once before the main data collection. Training is intended to not only identify potential problems but also help the country representatives to familiarize themselves with the procedures and software. In addition, data entry manuals that are based on the IEA technical standards (Martin et al. 1999) supplement this training and provide an ongoing source of guidance for data entry. In these “train the trainer” workshops, mock materials are used in hands-on practice sessions and these are made available also for further in-country training of data entry staff that are organized subsequently by country representatives.

In order to meet the tight project timelines and to identify systematic data capture errors as early as possible, data capture starts during the data collection period. All sources of data, including the questionnaires and test booklets used, need to be stored in an orderly fashion to permit countries to consult original materials during data processing and verification procedures at the IEA. Any data inconsistencies that may arise (e.g., number of years worked at the school is higher than number of years worked in total) have to be manually verified to assure that they are not a result of mistakes during the data entry stage.

9.2.1 Data Capture from Paper-Based Instruments

Each country receives international versions of survey instruments and corresponding codebooks for data capture. When countries choose to or need to adapt some of the survey instruments to reflect national conventions, they need to reflect this in the

codebooks before entering data. To ensure that the codebooks follow the structure of the survey instruments, a test data capture of each survey instrument has to be done. Once the correct adaptation of the codebooks is verified, data entry can start.

The training provided and the common rules for manual data capture ensure that data is entered comparably by all personnel who are responsible for data entry. Data must be recorded exactly as listed in the survey instrument, no interpretation is allowed. When responses are omitted or cannot be interpreted, study-specific missing values have to be entered.

9.2.2 Software Used for Data Capture

IEA developed the Data Management Expert (DME) to improve the quality of data entry and standardize data capture procedures. The DME evolved from the previous Windows Data Entry Manager (WinDEM) system. The underlying structure file (codebook) ensures that all entered data is verified against all possible valid and missing values. Wild codes (codes that are not defined as valid or missing) or out-of-range values can either not be entered at all or have to be confirmed before they can be entered. At the beginning of the data processing, DME additionally checks that each record is entered only once, which decreases the chance of duplicate IDs. It also allows the adaptation of the structure files to account for any national deviations from the international survey instrument structure (national adaptation). In comparison to WinDEM, where multiple codebooks were needed and no checking between files was possible, the DME codebook can include all survey instruments and thus allows for additional consistency checks between these (e.g., it can be used to check if a student completed a context questionnaire but no data for a test booklet was entered). The checks are written in SQL (structured query language) and can be customized for each study.

9.2.3 Quality Control: Data Entry

To ensure data capture personnel are properly trained and adherence to the data capture rules, a certain proportion of survey instruments have to be entered twice; this is called “double punching.”¹ A minimum agreement of 99% has to be reached for data to be accepted for submission. Countries are encouraged to enter data from paper survey instruments as early as possible during the data collection period so that possible systematic misunderstandings or mishandlings of data-capture rules can be identified quickly and appropriate remedial actions initiated, for example, further national center staff training.

¹The term “double punching” dates from a time when physical punch cards were used to enter data; two cards were punched from the same data and were visually compared for any differences.

During and after data capture of paper survey instruments national center staff are required to run a number of checks using the IEA data capture software. This ensures that data submitted to the IEA for further processing fulfill the initial quality requirements.

The following checks are included in the DME software and NRCs are asked to run them regularly:

- Unique ID check: a unique ID check ensures that data for each respondent's questionnaire is entered only once;
- Validation check: a check validating all entered data against the structure files (codebooks); and
- Record consistency check: a number of study specific checks verifying data across different survey instruments.

Most countries capture their data manually using trained data entry personnel. When countries choose to scan their paper instruments, they are required to provide proof of their scanning reliability by scanning instruments twice and comparing the output. This corresponds to the double punching of manual data capture.

In addition, the IEA Windows Within School Sampling Software (WinW3S), which operates the participation tracking database, offers another set of checks that must be run before data is submitted. Depending on the study design, all or a subset of the following checks are available and must be undertaken. WinW3S allows NRCs to check whether:

- The data is available in a different administration mode than that currently entered in the WinW3S database (online versus paper administration);
- Participation, exclusion, or questionnaire return status in the WinW3S database matches the data availability in the DME database, the Online SurveySystem data tables, or the eAssessment database;
- The teacher subject code in the WinW3S database is consistent with the Teacher Questionnaire data in the DME database or the Online SurveySystem data tables;
- The assigned Booklet ID in the WinW3S database differs from the booklet entered in the DME database;
- Data exists for a booklet, but this booklet has not been assigned to a student;
- There are inconsistencies between the booklets assigned for reliability scoring and data availability for these booklets.

9.3 Scoring Cognitive Data: Test Booklets

The cognitive (achievement) test booklets consist of both multiple choice and constructed-response items. In order to allow testing of a larger number of items, they are usually grouped into different blocks that are then rotated in the test booklets.

Multiple-choice items are questions in which respondents are asked to select the correct answer from a list of different, often similar answers. Multiple-choice items can be machine scored and do not need to be evaluated by trained scorers.

Constructed-response items are questions that require students to provide a written answer, give a numerical result, complete a table, or provide a drawing. At the national centers, scorers trained by the scorers that attended the IEA international scorer training evaluate responses to these questions based on scoring guidelines provided by the international study center (ISC), which identify specific criteria for assigning a particular score. All country representatives are required to attend a scorer training session prior to the field trial and main data collection. At these training sessions all items, possible student answers, and different scoring codes are discussed with attendees. This ensures a common understanding and interpretation of the scoring guidelines.

9.3.1 Process of Scoring Constructed-Response Cognitive Items

The scoring teams within countries are divided into two groups, each with one professionally trained supervisor. The scoring supervisor moderates and answers all questions from scorers and reads a sample of scored responses to monitor the scoring reliability. This process is also called “back-reading” and essential for identifying scorers who do not understand particular scoring instructions. In such cases, scorers may need to be retrained or replaced when necessary.

A critical component of the scoring procedures is monitoring the quality of the scoring and calculating inter-rater reliabilities both within cycle and across cycles for linkage. In each cycle of any IEA international large-scale assessment, a certain study specific amount (ranging between 15 and 35%) of these items have to be scored by two different scorers. To simplify the administration process of reliability scoring of paper instruments, whole booklets (instead of single items) are randomly selected by the software that is also used to assign booklets to the students (IEA WinW3S). Within these booklets all constructed-response items are reliability scored. In order for the reliability scoring to be blind, the reliability scoring is completed first, with scores recorded on a separate scoring sheet and not in the booklets. The main scoring is completed after that, with scores entered in predefined fields in the booklets directly (see Table 9.2).

Table 9.2 Scoring responsibilities

Step	Team A	Team B
1	Scores reliability booklets from set B Records scores on separate reliability scoring sheets	Scores reliability booklets from set A Records scores on separate reliability scoring sheets
2	Scores all booklets from set A Records scores directly into the booklets	Scores all booklets from set B Records scores directly into the booklets

When scoring items that are available electronically, the IEA Coding Expert will display items without any previously assigned score. The assignment of items for reliability scoring does not have to be connected to whole booklets. The reliability items will be divided equally between two groups of scorers (team A and B) into set A and B. While scorer group A scores the items of set B, scorer group B scores the items of set A. When they are done, the sets of items are exchanged and scored a second time by the other group of scorers.

9.3.2 Software Used for Scoring Data

IEA has developed different software solutions to support the scoring process. The data capture software DME (and formerly WinDEM) provides two check procedures that help countries to assess the reliability of their scoring. One procedure checks if the same items for the same student are scored by two different scorers, while another compares the scores of the main scorer with that of the reliability scorer and provides the user with the agreement rate between the scorers. These two checks provide NRCs with information on two major quality requirements. Countries are urged to perform the checks continuously to identify as early as possible in the process whether any consistent misunderstandings exist among scorers. Should this happen, further training or replacement of scorers is necessary.

The IEA Coding Expert undertakes the same checks as the IEA DME application but offers additional features. When items are electronically available (e.g., scanned or imported from any CBA system), they can be displayed directly using the IEA Coding Expert. In this case, scorers enter their scores directly into the software and no additional manual data capture of scores is necessary. The IEA Coding Expert offers a scorer management system that creates unique logins for each scorer and assigns items for main and reliability scoring automatically to each scorer. Should scorers be trained to score only specific items, the IEA Coding Expert also considers this and assigns items accordingly. This ensures that the main scoring and reliability scoring of items are not undertaken by a sole scorer and also enables the scoring supervisor to monitor agreement rates in real time.

Before the development of the IEA Coding Expert, the IEA provided countries with separate software for the cross-country scoring reliability study (CCSRS) and the trend scoring reliability study (TSRS). These two software packages covered the same functionalities as the IEA Coding Expert. With the development of the IEA Coding Expert they became obsolete, since a single software package that can be used for all scoring tasks is more convenient.

Some constructed response items in the eTIMSS 2019 study were machine scored using SQL scripts. The machine scoring was undertaken after students completed the test, and not during the test.

9.3.3 *Quality Control*

The scoring quality is measured as the agreement rate between the main and reliability scorer (inter-rater reliability). An occasional error, or an understandable disagreement when the rules are not sufficiently clear is part of a judgmental scoring process and expected. However, consistent errors in categorizations arising from lack of understanding about the intent of the scoring guide are more serious. In such cases, all of the concerned constructed-response items need to be checked and corrected, not just the items selected for reliability scoring. If the error in understanding can be narrowed down to specific scorers, only items scored by those scorers need to be checked. Naturally, the goal is to have 100% or perfect agreement among scorers. An agreement between scorers above 85% is considered good and agreement above 70% is considered acceptable. Percentages of agreement below 70% are a cause for concern. The finally achieved agreement rate is also used during data adjudication, namely when reviewing data quality and making decisions about annotations for the reporting of data.

To ensure that there is continuity of scoring of the same items between the cycles of TIMSS and PIRLS, the TSRS has to be conducted by all countries that are participating in the study cycle that also participated in the previous cycle. The TSRS allows scorers of the current study cycle to score student responses collected during the previous cycle and is conducted using the IEA Coding Expert software. The responses are scanned by IEA following the previous cycle of the assessment.

All scorers who participate in scoring of the items of the current study cycle have to participate in the TSRS. Similar to the within-country reliability scoring, the TSRS blends with the main scoring procedure and is ongoing throughout the scoring process.

To verify consistent scoring between countries in the TIMSS and PIRLS studies, participating countries also need to perform a CCSRS. This gives an indication of how reliable the scoring is done across countries. Student responses included in the CCSRS are those related to items collected from English-speaking countries during the administration of the previous cycle of the assessment. Just like the TSRS, the CCSRS is conducted using the IEA Coding Expert software. The same set of student responses in English will be scored by all participating countries.

All scorers who participate in scoring of the items of the current study cycle should participate in the CCSRS. Similar to the TSRS, the CCSRS blends in with the main scoring procedure and is ongoing throughout the scoring process.

9.4 Coding Data

9.4.1 *Process of Coding Data*

Some IEA studies collect data on students' parental occupations using an open-ended text format. Before they can be internationally compared, they need to be transferred to an internationally comparable numerical code. For the coding of occupational data, IEA studies use a framework recommended by the International Labor Organization (ILO). The International Standard Classification of Occupations, ISCO-08 (ILO 2012) is a broadly accepted and accessible international classification of occupational data and a revised and improved version of its predecessor, ISCO-88, which was used in in previous study cycles.

Countries are encouraged to hire people with prior experience in this area of the coding. The coding team within countries is led by one supervisor. Supervisors review aggregated codes carefully and monitor coding progress. To further ensure high coding quality and common understanding of the coding rules, a study specific amount (ranging between 10 and 15%) of questions are coded twice.

Coding quality is proportional to the quality of student responses. The more detailed the information that students provide on their parents' occupations, the easier it is for coders to assign correct codes. Therefore, countries are advised to make administrators aware of the content of the questions that students will have to answer. If possible, students should be informed prior to the survey that these questions regarding their parents' occupation will be asked so that they have time to prepare their responses; if needed, schools should also explain to parents that they will need to help their children by providing the information that enables them to answer these questions. This would reduce the number of vague and omitted responses provided by students.

9.4.2 *Software Used for Coding Data*

To support the coding process, different software solutions have been developed by the IEA. The IEA data capture software DME supports two possible scenarios when coding questions. When questions are coded directly on the paper questionnaires and reliability coding sheets (occupation double coding sheets), the final ISCO-08 codes are entered directly into the codebooks with DME software. The software provides two check procedures that support countries in determining their reliability of the coding. One check procedure compares the values of the main and reliability codes and provides the user with the agreement rate between the coders. Another checks if the same question for the same student is coded by two different coders as required. These two checks provide NRCs with information on two major quality requirements at any time in the process.

Alternatively, it is also possible to enter the text responses from the occupational coding questions into the DME software. Once all data has been entered this way, the occupation data can be extracted to a specially designed Excel file that can be used for coding. The advantage of this procedure is that the Microsoft Excel file can be sorted by job title, which will make coding faster and more reliable. In the Microsoft Excel file the same check procedures are included as in the DME software.

Countries are urged to perform the checks continuously to identify any consistent misunderstandings among scorers as early as possible in the process and to initiate appropriate remedial actions when necessary, such as staff retraining.

The IEA Coding Expert covers the same checks as the IEA DME and offers additional features. When items are electronically available (e.g., scanned or imported from any CBA system) they can be displayed directly through the IEA Coding Expert. The items are coded in the software and no additional data capture of codes is necessary. The IEA Coding Expert offers a coder management system that creates unique logins for each coder and assigns questions for main and reliability coding automatically to each coder. This ensures that the same coder does not code both the main and reliability samples and it also enables the coding supervisor to monitor agreement rates in real time.

9.4.3 *Quality Control*

The coding quality is measured as the agreement rate between the main coder and reliability coder. An occasional error or an understandable disagreement when the rules are not sufficiently clear is part of a judgmental coding process. This is all part of the information about the coding reliability that is entered into the database. Consistent errors in classification of responses to occupational questions arise from misunderstandings. As for the other data coding activities, the goal is to have 100% or perfect agreement among coders. An agreement of more than 85% between coders is considered good and agreement above 70% is considered acceptable. Percentages of agreement below 70% are a cause for concern. The finally achieved agreement rate is also used during data adjudication, to review data quality and make decisions about annotations for reporting of occupational data. In total, about one sixth of the responses need to be double-coded. Reviews of coder discrepancies may indicate problems with the work of particular coders or a particular set of occupations that are difficult to allocate. In these cases, further training or the replacement of some coders may be advisable, thus IEA recommends coder agreement checks are undertaken at an early stage in the coding process. Where such data exist, it is also recommended that data are cross-checked against external/historical data sources.

9.5 International Data Processing

9.5.1 *Processes in International Data Processing*

The ultimate objective of data processing is to ensure the availability of consistent (reliable) and valid data for analysis. Once data collection within a country is completed and all data is available, participating countries submit their materials to the IEA. IEA then verify that all required materials that were sent are complete and fulfill all previously defined requirements. The receipt of materials is tracked in databases and confirmed to countries. In those cases where materials are missing, incomplete, or faulty, countries are contacted to resolve all issues. Materials are either sent via post, fax machine, email or uploaded to secure servers. Data collected during survey administration via IEA's servers (e.g., data from online context questionnaires) is already available and does not have to be resubmitted. If this is the case, countries have to confirm that data collection is finished and that access to online context questionnaires, the coding system, or the CBA system can be disabled. This ensures that no data is collected after the official data collection window and that data processing only starts when all materials are finalized. It remains, however, possible to reactivate access when requested by countries in agreement with the ISC to address data quality issues that may arise unexpectedly. This might be the case when there is reason to believe that additional time for data collection will improve low participation rates.

After all data are submitted to the IEA, final data processing commences. The objective of the process is to ensure that the data adheres to international formats, that information from respondents can be linked across different instrument files, and that the data accurately and consistently reflects the information collected within each participating country.

During the data processing, the IEA reports all inconsistencies that were found in the data back to countries to resolve all remaining issues. When all open issues regarding respondents' participation or reason for non-participation are resolved, the weighting of the sample data can start. Although the international sampling plan is prepared as a self-weighting design (whereby each individual ultimately has the same final estimation weight), the actual conditions in the field, non-response, and the coordination of multiple samples often make that ideal plan impossible to realize. To account for this, weights are therefore computed and added to the data.

Data submitted to the IEA come from a variety of different sources. Data for any individual respondent can come from any of the following: a tracking database, a database that is used for data entry and submission of paper questionnaires, a database that is used for data entry and submission of paper booklets, a database in which responses from the online questionnaires are stored, and a database in which responses from the online booklets are stored. The first step of the data processing, the data import, is to match-merge data from all sources. During this step, no specific checks are generated, but duplicate records are identified. Multiple data sources for a single respondent (i.e., those who responded to the same questionnaire on paper and

online, or those who responded to the same online questionnaires in two different languages) are flagged and need to be resolved. In most cases, the data pattern shows that a respondent changed their mind about their original choice of administration mode or language of questionnaire and very quickly abandoned their original choice to complete the questionnaire in another mode or language. In that case the data of the incomplete records are simply deleted. If, however, two equally complete data records with different data exist, the IEA asks the country to advise which record should be deleted. This issue of duplicate records is quite common for all studies, but usually affects only a few cases per country. Data processing can only continue with the next step once these issues of duplicate records are resolved.

In the next step after the data import, the structure of the national data is checked against the international data structure. The main goal of this is to document deviations from international standards both in data files and instruments, and to modify the national data according to the international design to ensure that the resulting data is internationally comparable. This is not only important for later analysis but also for all following cleaning steps, and ensures a consistent treatment of all data. National adaptations to the international survey instruments are agreed on and documented in special templates (national adaptation forms) before the data is collected. During the structure check, automated checks flag all deviations from the international format, which are then crosschecked against the adaptations documented in the national adaptation forms.

The main objective of cleaning is to identify any issues and deviations at the observation level. All deviations that appear for single observations or groups of respondents (e.g., students within one school) are reported. The data checks during the data processing can be divided into two major groups: ID and linkage cleaning, and background cleaning.

The automated checks of the ID and linkage cleaning compare the available data from the survey instruments with the reported available data in the tracking database. Checks ensure that the hierarchical ID system of the study was followed and thus a linkage between different respondent groups (e.g., students, parents, teachers, and principals) is possible, and that all tracking variables (e.g., student age and participation status) are assigned valid values and that the information in them is not contradictory to the actually available data from context questionnaires and test booklets.

The background cleaning checks verify the data in the context questionnaires. Depending on the item format certain data patterns can be expected. Answers to numerical questions are expected to fall within a certain range (e.g., student age), the sum of items of percentage questions (e.g., percentage of time the principal spends on the tasks in the school) is expected to equal 100%, the answers following a filter-dependent question are expected to be omitted when the filter question has been answered positively (e.g., does a school have a library vs. how many books are in the library), and the answer of logically dependent questions are expected to be consistent (e.g., enrollment of students in target grade is not expected to be greater than enrollment of students in the whole school).

The cleaning step produces findings for all checks. These findings can either be resolved with additionally available information (e.g., tracking forms or other supplementary documentation submitted by countries) or if the findings cannot be resolved directly, NRCs are contacted for further information and advice.

After all issues detected during the data processing steps are either resolved or confirmed by NRCs, the data processing commences with the next step, termed post cleaning. During post cleaning, the data undergo various major modifications.

To avoid discouraging respondents from answering questionnaires either incompletely or not at all, questions are never asked twice to the same respondent. For example, the teacher general background information would be the same when the mathematics teacher is also the science teacher. In those instances, data from overlapping questions between both questionnaires are copied over from one questionnaire to the other.

During post cleaning, a special missing code is assigned to questions that were deemed “not reached” to distinguish them from “omitted” responses. “Omitted” questions are those that a respondent most probably read, but either consciously decided not to answer or accidentally skipped; that is, the respondent started answering the questions but stopped answering before the end of the survey instrument, likely due to a lack of time, interest, or cooperation. “Not reached” responses are exclusively located towards the end of the survey instrument. To code as not reached, the last valid answer given in a survey instrument is identified. The first omitted response after this last answer is coded as omitted, but all following responses are then coded as not reached. Analyzing the frequency of the not reached code can give valuable information on the design of the survey instruments (e.g., length of context questionnaires, difficulty of test booklets). The not reached codes may be handled differently during the data analysis. In TIMSS, for example, the not reached items are treated as incorrect responses, except during the item calibration, where they are considered as not having been administered.

Not all processing steps necessarily lead to either automatic or manual recoding of data. Some data remained unchanged although a finding is reported for the affected records (e.g., contradictory/inconsistent filter usage). If a larger number of unresolved inconsistencies remains for a specific check, the data under consideration are carefully reviewed by IEA. For some cases, final actions or recodings during the post-cleaning phase, that is after all country processing feedback is taken into account, are agreed between the IEA and external stakeholders (such as NRCs and contractors).

In the final step, weights and scores that are computed based on processed data are merged with the processed data. All external data which is merged during the processing of the data undergoes its own rigorous checking procedures (see Chaps. 7 and 11).

The final check repeats all checks from the structure check and the cleaning checks. This is a quality measure to ensure that the data modifications during the post cleaning do not affect the data in any unintended way. However, the main goal of the final checks is to check the data that has been merged during the final merge.

The export step of the data processing produces data files in different formats. The data files produced during the export are frequently exchanged with partners, stakeholders, and countries. Data is available in various file formats (e.g., *.csv, *.sav, *.sas7bdat, *.dta) to facilitate further data analysis with different statistical software packages. Data can be exported at any time during the data processing period.

To protect the confidentiality of the respondents, certain disclosure avoidance measures are applied at the international level, which are consistent for all countries, and at the national level, which concern only specific national datasets. The most common measures across studies are scrambling of IDs and the removal of tracking or stratification variables. Measures at the national level can range from the removal of specific variables to the removal of complete datasets. Usually two versions of the international databases are created: a public-use file (PUF), available without any restrictions to any interested person, and a restricted-use file (RUF), available only upon special request by researchers. Researchers who want to use the RUF need to formally apply and this application is reviewed by IEA before access the restricted-use file is granted with requisite confidentiality rules. Unlike the PUF, the RUF includes confidential data (e.g., data identifying students' birth months and years).

9.5.2 Software Used for International Data Processing and Analysis

IEA uses three different systems to process all data from international studies. Regardless of the programs used, the processing steps and data checks are the same, and are system independent; differences are due to specific study designs and survey instruments. In theory, all three systems can be used for the data processing of all studies. While some are more powerful and convenient when processing large amounts of data, others are more easily adaptable by less technically-oriented staff. Finally, the timing of a study, the available budget, and the amount of data determine which cleaning program is used.

Originally, the data processing for all international studies was implemented using SAS (SAS Institute Inc. 2013). Due to the lack of readily available SAS programmers in many countries, IEA has also developed processing tools in SQL and SPSS (IBM Corp. 2017).

The international studies unit at IEA collects data processing requirements from all studies before they are implemented by programmers. This ensures that new developments between studies are exchanged and each new study benefits from the latest developments. When the data processing programs are developed, they are thoroughly tested using simulated data sets containing all the expected problems or inconsistencies. Providing data and programs in the different file formats maximizes the accessibility and utility of study data, further enhancing consequential validity.

9.5.3 *Quality Control*

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 those in charge, the data quality control process includes thorough testing of all data processing programs with simulation datasets containing all possible problems and inconsistencies. Deviations from the cleaning sequence are not possible, and scope for involuntary changes to the cleaning procedures is minimal. All systematic and individual data recodings are documented and provided to NRCs so that they can thoroughly review and correct any identified inconsistencies.

During the data processing, data is continuously exchanged with partners, countries, and any other stakeholders of the study. Before data updates are sent out, data are compared with previously sent data and any deviations are checked and verified. This ensures that only expected changes have been implemented in the data.

Univariate descriptive statistics are produced to help to review the content responses of the questionnaires. One file per sample and respondent level is created. Each presentation of univariate statistics provides the distribution of responses to each of the variables for each country.

9.6 Conclusions

Throughout the process of post-collection data capture, scoring, coding, and data processing, common quality control procedures ensure reliable data. These quality control procedures include a set of study-specific rules that all participating countries have to adhere to and customized software products that support both the NRCs and IEA in checking the adherence to these rules. This ensures that data collected by multiple countries, in multiple languages, and from respondents at different levels can be linked within countries and compared across countries.

At the conclusion of the study, IEA creates an international database; while most data is publicly available (the PUF), and the remainder is available only on request for formally approved research uses (the RUF).

References

- Fishbein, B., Martin, M., Mullis, I.V.S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6, 11. Retrieved from <https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-018-0064-z>.
- IBM Corporation. (2017). *IBM SPSS Statistics for Windows (Version 25.0)* [Computer software]. Armonk, NY: IBM Corp. Retrieved from <https://www.ibm.com/analytics/spss-statistics-software>.

- ILO. (2012). *International Standard Classification of Occupations ISCO-08. Volume 1: Structure, group definitions and correspondence tables*. Geneva, Switzerland: International Labour Office. Retrieved from <https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf>.
- Martin, M. O. Rust, K., & Adams, R. J. (Eds.). (1999). *Technical standards for IEA studies*. Amsterdam, the Netherlands: IEA. Retrieved from <https://www.iea.nl/publications/iea-reference/technical-standards-iea-studies>.
- SAS Institute Inc. (2013). *SAS university edition 9.4 [Computer software]*. Cary, NC: SAS Institute Inc. Retrieved from https://www.sas.com/en_us/software/university-edition.html.

Alena Becker is co-head of the International Studies Unit coordinating the data management activities for all international studies unit at IEA. Her position also includes supervising the international study centers at IEA Hamburg.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.