10

Data Management Procedures

INTRODUCTION

The PISA assessment establishes standard data collection requirements that are common to all PISA participants. Test instruments include the same test items in all participating countries, and data collection procedures are applied in a common and consistent way amongst all participants to help ensure data quality. Test development is described in Chapter 2, and the data collection procedures are described in this chapter.

As well as the common test elements and data management procedures, the opportunity also exists for participants to adapt certain questions or procedures to suit local circumstances, and to add optional components that are unique to a particular national context. To accommodate the need for such national customisation, PISA procedures need to ensure that national adaptations are approved by the Consortium, are accurately recorded, and where necessary the mechanisms for re-coding data from national versions to a common international format are clearly established. The procedures for adapting the international test materials to national contexts are described in Chapter 2 and the procedures for adapting the questionnaires are described in Chapter 3. The mechanisms for re-coding data from national versions to a common international format are described in this chapter.

As well as planned variations in the data collected at the national level, the possibility exists for unplanned and unintended variations finding their way into the instruments. Data prepared by national data teams can be corrupted or inaccurate as a result of a number of unintended sources of error. PISA data management procedures are designed to minimise the likelihood of errors occurring, to identify instances where errors may have occurred, and to correct such errors wherever it is possible to do so before the data are finalised. The easiest way to deal with ambiguous or incorrect data would be to delete the whole data record containing values that may be incorrect. However, this should be avoided where possible since the deleted data records results in a decrease in the country's response rate. This chapter will therefore also describe those aspects of data management that are directed at identifying and correcting errors. These procedures applied for both the pencil and paper and computer-delivered components of PISA 2009.

The complex relationship between data management and other parts of the project such as development of source materials, instrument adaptation and verification, as well as school sampling are illustrated in Figure 10.1. Some of these functions are located within national centres, some are located within the international Consortium, and some are negotiated between the two.

Data management procedures must be shaped to suit the particular cognitive test instruments and background questionnaire instruments used in each participating country. Hence the source materials provided by the Consortium, the national adaptation of those instruments, and the international verification of national versions of all instruments must all be reflected in the data management procedures. Data management procedures must also be informed by the outcomes of PISA sampling procedures. The procedures must reliably link data to the students from whom they came. Finally, the test operational procedures that are implemented by each national centre, and in each test administration session, must be directly related to the data management procedures.

■ Figure 10.1 ■

Data management in relation to other parts of PISA

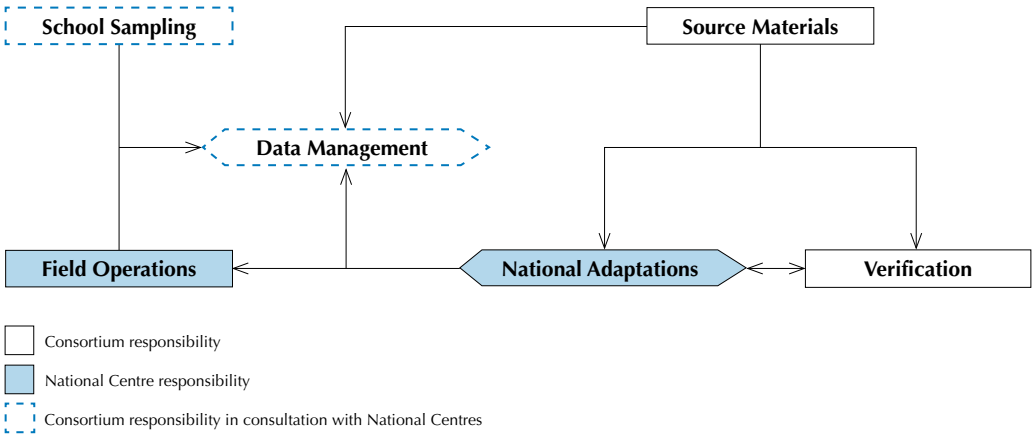


Figure 10.2 illustrates the sequence of major data management tasks in PISA, and shows something of the division of responsibilities between national centres, the Consortium, and those tasks that involve negotiation between the two. This section briefly introduces each of the tasks. More details are provided in the following sections.

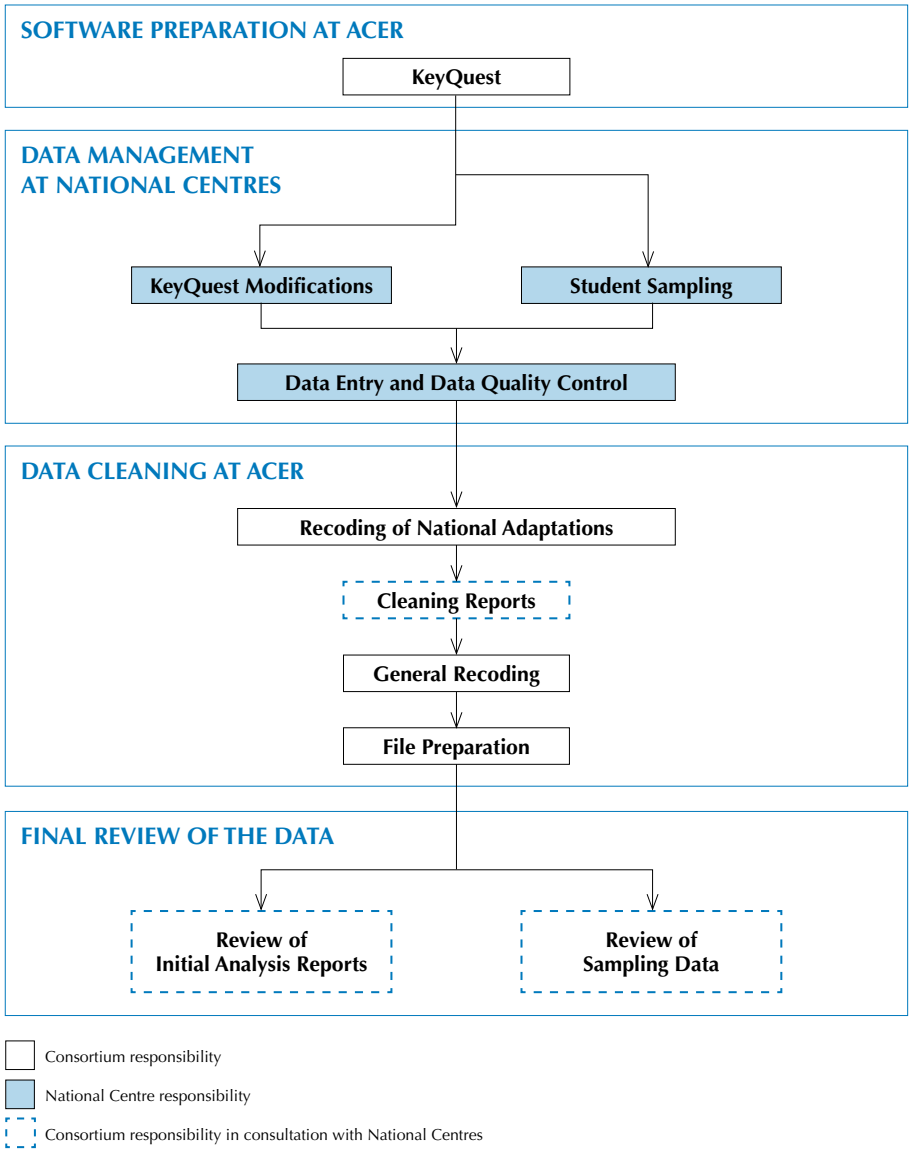
First, ACER provides the data management software *KeyQuest* to all national centres. *KeyQuest* is generic software that can be configured to meet a variety of data entry requirements. In addition to its generic features, the latest version of *KeyQuest* was pre-configured specifically for PISA 2009.

KeyQuest was preconfigured with all the PISA 2009 standard instruments: cognitive test booklets, background and contextual questionnaires, and student tracking instruments that are derived following implementation of the school sampling procedures. However, it also allows for instrument modifications such as addition of national questions, deletion of some questions and modification of some questions. A prerequisite for national modification of *KeyQuest* is Consortium approval of proposed national adaptations.

After the national centres receive *KeyQuest*, they carry out student sampling and they implement *KeyQuest* modifications as a part of preparation for testing. By that time the variations from the core PISA sampling procedures such as national and international options (see Chapter 6) and the proposed national adaptations of the international source instruments (see Chapters 3 and 6) were agreed with Consortium and all national versions of instruments have been verified.

■ Figure 10.2 ■

Major data management stages in PISA



Following test administration and coding of student responses, national centres are required to enter the data into *KeyQuest*, to perform validity reports to verify data entry, and to submit the data to ACER.

As soon as data are submitted to ACER, additional checks are applied. During the process of data cleaning, ACER sends cleaning reports containing the results of the checking procedures to national centres, and asks national centres to clarify any inconsistencies in their database. In the questionnaires for example such inconsistencies might include the number of qualified teachers in a school exceeding the total number of teachers or unlikely (though not impossible) situations such as parents with higher degrees but no secondary education. The national data sets are then continuously updated according to the information provided by the national centres. The cleaning reports are described in more detail below.

Once ACER has received all cleaning reports from the national centres and has introduced into the database all corrections recommended in these reports, a number of general rules are applied to the small number of unresolved inconsistencies in the PISA database.

At the final data cleaning stage national centres are sent the initial analysis reports containing cognitive test item information and frequency reports for the contextual questionnaires. The national centres are required to review these reports and inform ACER of any inconsistencies remaining in the data. Further recodings are made after the requests from the national centres are reviewed. At the same time sampling and tracking data is sent to Westat, analysed and when required further recodings are requested by Westat and implemented at ACER. At that stage the database is regarded as final, and is ready for submission to the OECD.

DATA MANAGEMENT AT THE NATIONAL CENTRE

National modifications to the database

PISA's aim is to generate comparable international data from all participating countries, based on a common set of test instruments. However, it is an international study that includes countries with widely differing educational systems and cultural particularities. Due to this diversity, some instrument adaptation is required. Hence verification by the Consortium of national adaptations is crucial (see Chapter 3). After adaptations to the international PISA instruments are agreed upon, the corresponding modifications in *KeyQuest* are made by national centres.

Student sampling with *KeyQuest*

Parallel to the adaptation process national centres sample students using KeyQuest. The student sampling functionality of *KeyQuest* was especially developed for the PISA project. It uses a systematic sampling procedure by computing a sampling interval. *KeyQuest* samples students from the information in the list of schools. It automatically generates the student tracking form (STF) and assigns one of the rotated forms of test booklets to each sampled student. In the process of sampling, *KeyQuest* uses the study programme table (see Chapter 3), and the sampling form designed for *KeyQuest* (SFKQ, see Chapter 4) which were agreed with the National Centres via MyPISA and imported into *KeyQuest*.

The student tracking form and the list of schools are central instruments, because they contain the information used in computing weights, exclusion rates, and participation rates. Other tracking instruments used in *KeyQuest* included the session report form which is used to identify the language of test for each student. The date of the testing session that the student attended obtained from the session report is used in conjunction with the date of birth of the student from the tracking form to calculate the age of the student at the time of testing.

Data entry quality control

The national adaptation and student sampling tasks are performed by staff at each national centre before testing. After testing the data entry and the validity reports are carried out by the national centres.

Validation rules

During data entry *KeyQuest* captures some data entry errors through the use of validation rules that restrict the range and type of values that can be entered for certain fields. For example, for a standard multiple-choice item with four choices, one of the values of 1-4 each corresponding to one of the choices (A-D) that is circled by the student can be entered. In addition, code 9 was used if none of the choices was circled and code 8 if two or more choices were circled. Finally code 7 was reserved for the cases when due to poor printing an item presented to a student was illegible, and therefore the student did not have access to the item. No other codes could be entered.

Key violations

Further, *KeyQuest* was programmed to prevent key violations. That is, *KeyQuest* was programmed to prevent the duplication of so called keys, which are usually the combination of identifier codes. For example, a data record with the same combination of stratum and school identifiers could not be entered twice in the school questionnaire instrument.

KeyQuest also allows double entry of the test and questionnaire data and monitoring of the data entry operators. These procedures are described below.

Monitoring of the data entry operators

The data entry efficiency report was designed specifically for PISA 2009 to keep the count of records entered by each data entry operator and the time required to enter them. The Consortium recommended to all countries to use some part of these procedures to assure quality of the data entry.

Double coding of occupational data

Another optional procedure for PISA 2009 was the double coding of occupational data. The double coding allowed national centres a check of the validity of the data and it allowed identification of the areas where supplementary coding tools could be improved. The main coding tool was the *ISCO Manual* (ILO, 1990) with the small number of additional codes described in the PISA 2009 *Data Management Manual*.[1] The supplementary coding tools would typically include coding instructions, a coding index, and training materials developed at the national centre.

Under this procedure the occupational data from the student questionnaires and parent questionnaires (if applicable) were coded twice by different coders and entered into two *KeyQuest* tables specifically designed for this purpose. Then the double entry discrepancies report was generated. The records for which there were differences between ISCO Codes entered into the two tables were printed on the report, analysed by the data manager and acted upon. The possible actions would be improvement of the instructions if the same error was systematically produced by different coders, and/or further training of coders that were making more errors than others. Finally, the Consortium expected all discrepancies printed on the report to be resolved before the data were submitted to ACER.

The national centres that participated in this option commented on the usefulness of the procedures for training of the coding staff. The possibilities for analysis by the Consortium of the data from this option were limited due to the language constraints. One of the results was that those countries that required their coders to enter a word description as well as a four-digit code had fewer discrepancies than those that required only a four-digit code. This led to a reinforcement of the ILO recommendation that procedures should involve entering occupation descriptions first and then coding them, rather than coding directly from the questionnaires.

Validity reports

After the data entry was completed the national centres were required to generate validity reports from *KeyQuest* and to resolve discrepancies listed on these reports before submitting data to ACER.

The structure of the validity reports is illustrated by Figure 10.3. They include:

- comparison between tracking instruments and sampling verification (tracking instruments, sampling verification);
- data verification within tracking instruments (tracking instruments specific checks);
- comparison of the questionnaire and tracking data (student questionnaire-student tracking form specific checks, identity checks questionnaires, identity checks occupation);
- comparison of the identification variables in the test data (identity checks booklets, identity checks DRA); and
- verification of the reliability data (reliability checks).

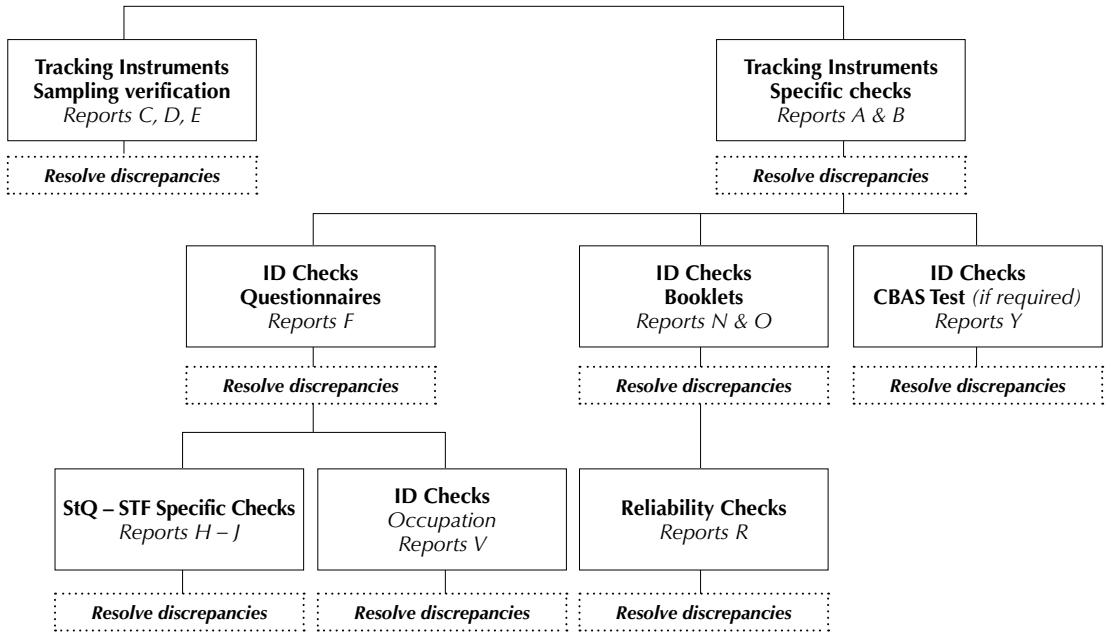
Some validity reports listed only incorrect records (e.g. students whose data were entered in more than one booklet instrument), whilst others listed both incorrect and *suspicious* records, which were records that could have been either correct or incorrect, but were deemed to be in need of confirmation. The resolution of discrepancies involved the following steps:

- correction of all incorrect records: e.g. students entered as "non participant", "transferred out of school" but who were also indicated on the student tracking form as having been tested; and
- an explanation for ACER as to how records on the report that were listed as suspicious, but were actually correct, occurred (e.g. students with special education needs were not excluded because it is the policy of the school).

Due to the complexity and significant number of the validity reports, a validity report checklist was designed. More details about the validity reports can be found in the PISA 2009 *Data Management Manual*.[2]

■ Figure 10.3 ■

Validity reports – general hierarchy



DATA CLEANING AT ACER

Recoding of national adaptations

When data submitted by national centres arrived at ACER, the first step was to check the consistency of the database structure with the international database structure. An automated procedure was developed for this purpose. For each instrument the procedure identified deleted variables, added variables and variables for which the validation rules had been changed. This report was then compared with the information provided by the NPM in the various adaptation spreadsheets such as the questionnaire adaptation sheet (see Chapter 3). For example, if a variable had been added to a questionnaire, the questionnaire adaptation sheet was checked by Core B to find out whether this national variable required recoding into the corresponding international one, or had to be set aside as being for purely national use and returned to the country. Once all deviations were checked, Core B sent necessary recodes for the submitted data to ACER to fit the international structure. All additional or modified variables were set aside and returned to the national centres in a separate file so that countries could use these data for their own purposes, but they were not included in the international database.

Data cleaning organisation

The data files submitted by national centres often needed specific data cleaning or recoding procedures, or at least adaptation of standard data cleaning procedures. To reach the high quality requirements, the Consortium implemented dual independent processing; that is, two equivalent processing tools were developed – one in SPSS and one in SAS – and then used by two independent data cleaners for each dataset.

For each national centre's data two analysts independently cleaned all submitted data files, one analyst using the *SAS*® procedures, the other analyst using the *SPSS*® procedures. The results were compared at each data cleaning step for each national centre. The cleaning step was considered complete for a national centre if the recoded datasets were identical.

DRA data

For countries which participated in the Digital Reading Assessment, the data file constructed from the DRA test delivery and online coding systems was introduced into the cleaning at the stage of processing with SAS and SPSS. A check on

student IDs was made with the cognitive data from *KeyQuest* and the DRA data was retained only for those students who had participated in the paper-based PISA assessment.

Cleaning reports

During the process of data cleaning, ACER progressively sent cleaning reports containing the results of the checking procedures to national centres, and asked national centres to clarify any inconsistencies in their database. The national data sets were then continuously updated according to the information provided by the national centre.

Many of the cleaning reports were designed to double check the validity reports, and if the data had been cleaned properly at the national centre, the cleaning reports would either not contain any records or would have only records that had been already explained on the validity reports. These cleaning reports were sent only to those countries whose data required additional cleaning.

However there were checks that could not be applied automatically at the national centre. For example, inconsistencies within the questionnaires could be checked only after the questionnaire data had been recoded back into the international format at ACER. These cleaning reports were sent to all national centres.

General recodings

After ACER received all cleaning reports from the national centres and introduced into the database all corrections recommended in these reports, the Consortium applied the following general rules to the unresolved inconsistencies in the PISA database (this was usually a very small number of cases and/or variables per country, if any):

- Unresolved inconsistencies regarding student and school identification led to the deletion of the record in the database.
- The data of an unresolved systematic error for a particular cognitive item was replaced by the not applicable code. For instance, if a country informed ACER about a mistranslation or misprint for an item in the national version of a cognitive booklet then the data for this item were recoded as *Not Applicable* and were not used in the subsequent analyses.
- If the country deleted a variable in the questionnaire, it was replaced by the not applicable code.
- If the country changed a variable in the questionnaire in such a way that it could not be recoded into the international format, the international variable was replaced by the not applicable code.
- All added or modified questionnaire variables were set aside in a separate file and returned to countries so that countries would be able to use these data for their own purposes.

FINAL REVIEW OF THE DATA

As an outcome of the initial data cleaning at ACER, cognitive, questionnaire, and tracking data files were prepared for delivery to the OECD and for use in the subsequent analysis by national centres and internationally.

Review of the test and questionnaire data

The final data cleaning stage of the test and questionnaire data was based on the data analyses between and within countries. After implementation of the corrections made on the cleaning reports and general recodings, ACER sends initial analysis reports to every country, containing information about their test and questionnaire items, with an explanation of how to review these reports. For test items the results of this initial analysis are summarised in six reports that are described in Chapter 9. For the questionnaires the reports contained descriptive statistics on every item in the questionnaire.

After review of these initial analysis reports, the NPM should provide information to ACER about test items that appear to have behaved in an unacceptable way (these are often referred to as dodgy items) and any ambiguous data remaining in the questionnaires. Further recoding of ambiguous data followed. For example, if an ambiguity was due to printing errors or translation errors a not applicable code was applied to the item.

Recoding required as a result of the initial analysis of international test and questionnaire data were introduced into international data files by ACER.

Review of the sampling data

The final data cleaning step of the sampling and tracking data was based on the analyses of tracking files. The tracking files were sent routinely country by country to Westat, the Consortium partner responsible for all matters related to sampling. Westat analysed the sampling and tracking data, checked it and if required requested further recodings, which were implemented at ACER. For example, when a school was regarded as a non-participant because fewer than 25% of students from this school participated in the test, then all students from this school were deleted from the international database. Another example would be a school that was tested outside the permitted test window. All data for students from such a school would also be deleted.

NEXT STEPS IN PREPARING THE INTERNATIONAL DATABASE

When all data management procedures described in this chapter were complete, the database was ready for the next steps in preparing the public international database. Student weights and replicated weights were created as described in Chapter 8. Questionnaire indices were computed or scaled as described in Chapter 16. Cognitive item responses were scaled to obtain international item parameters that were used to draw plausible values as student ability estimates (see Chapters 9 and 12).

Notes

1. For example, codes suggested by Ganzeboom & Treiman (1996) for very broad categories that sometimes appear in respondents' self-descriptions as well as in the cruder national classifications were used in PISA in addition to the standard ILO codes. These are: (1240) "Office managers", (7510) "Non-farm manual foremen and supervisors", (7520) "Skilled workers/artisans", (7530) "Apprentices", and (8400) "Semi-skilled workers". Another example is additional auxiliary codes that were later recoded as missing. These codes were: 9501 for home duties, 9502 for student, 9503 for social beneficiary (e.g. unemployed, retired, etc.), 9504 for "I don't know" and similar responses, and 9505 for vague responses.

2. Available at *www.pisa.oecd.org > what PISA produces > PISA 2009 > PISA 2009 manuals and guidelines*.