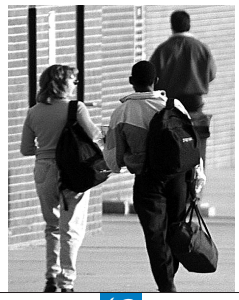


10

Data Management Procedures

INTRODUCTION

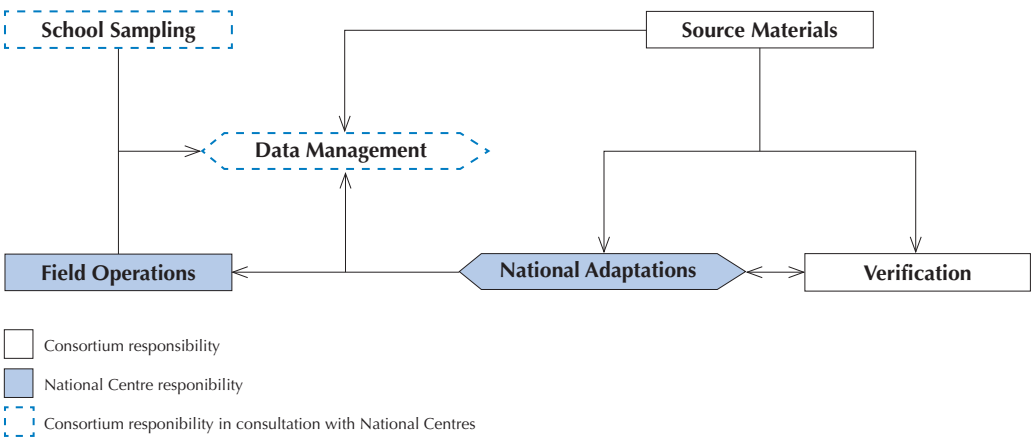
The PISA assessment establishes certain data collection requirements that are common to all PISA participants. Test instruments include the same test items in all participating countries, and data collection procedures are applied in a common and consistent way amongst all participants to help ensure 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 two 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 record containing values that may be incorrect. However, this should be avoided where possible since the deleted 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.

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.

Figure 10.1
Data management in relation to other parts of PISA



PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

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.

In summary, the data management must ensure that each student taking the PISA test is known, that the particular questions to which each student responds are known, and that the data generated by each student are the most accurate reflection possible of the responses provided by the student, and end up in the right cells of the final database.

Figure 10.1 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 2006.

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 Chapter 3 and Chapter 6) were agreed with consortium and all national versions of instruments have been verified.

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.

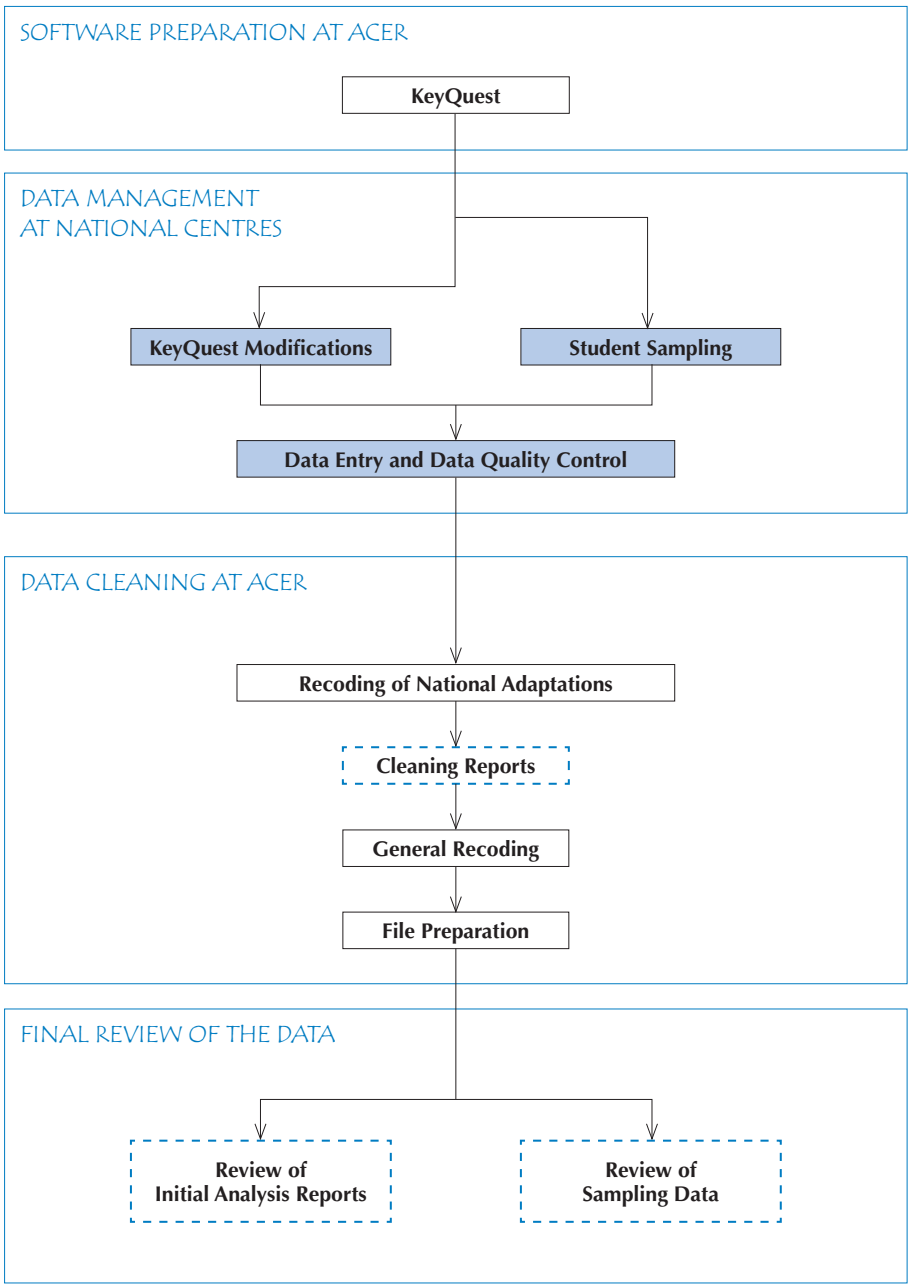
As soon as data are submitted to ACER, additional checks are applied. During the process of data cleaning, ACER sends cleaning reports containing the results of the checking procedures to national centres, and asks national centres to clarify any inconsistencies in their database. The national data sets 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.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Figure 10.2
Major data management stages in PISA



PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

KEYQUEST

KeyQuest is PISA's overarching data management tool. It is data management software that is dispatched to national centres before testing. *KeyQuest* is designed to facilitate student sampling, data entry and data validation.

KeyQuest was preconfigured with all the PISA 2006 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.

KeyQuest produces error messages when information is entered that violates its data validation rules, and it also generates validity reports. Validity reports list inconsistencies within the data and national centres are required to resolve these inconsistencies before they submit the data to ACER. In addition, the optional procedures for double entry of data and double coding of occupational data were developed and implemented by some national centres.

The use of the various *KeyQuest* functions by national centres is described in the next section.

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 (SPT, *see* Chapter 3), and the sampling form designed for *KeyQuest* (SFKQ, *see* Chapter 4) verified during adaptations 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 session report form together with the student tracking form are also used to calculate student age 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.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

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 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 2006 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 (as appropriate) to assure quality of the data entry.

Double entry facilities

In addition to that, the consortium recommended that at least 10% of the data was entered twice to assess the quality of the data entry. The *KeyQuest* double entry discrepancies report was designed to detect data entry errors by comparing data entered by different data entry operators. It was based on the assumption that the same random data entry error is unlikely to appear simultaneously. And therefore most data entry errors would be identified as a discrepancy between two parallel sets of data entered by different data entry operators.

Nine countries participated in a double data entry option that was included as part of the PISA 2006 field trial, which took place in 2005. In the participating countries double data entry was implemented for booklets 5 and 11. The index used to indicate the number of discrepancies was computed as follows:

10.1

$$D = \frac{\text{Number of discrepancies}}{\text{Number of strokes per student x Number of students}} \text{ x } 100\%$$

and the results are shown in Table 10.1.

While there was considerable variation between countries, the rate of discrepancies in all of the participating countries was low. The worst result was a discrepancy rate of 1.35% including both cognitive and attitudinal items. *KeyQuest* validation rules restricted the possibility of errors. This explains the low level of discrepancies.

Further to this analysis a simulation study was conducted that showed that the use of *KeyQuest* ensured the level of data entry errors was sufficiently low not to influence estimates of student achievement. In particular, the simulation study showed that if the percentage of discrepancies is lower than 4 percent, neither mean achievement nor standard errors of the means are changed significantly. For comparison the largest number of discrepancies in the real data from the double data entry option was 1.35% (see Table 10.1, country E).

Table 10.1
Double entry discrepancies per country: field trial data

	Number of students		Number of discrepancies		D	
	Booklet					
Country	5	11	5	11	5	11
A	125	118	132	20	0.66%	0.13%
B	131	134	107	40	0.51%	0.23%
C	166	169	178	223	0.67%	1.03%
D	92	102	3	33	0.02%	0.25%
E	100	101	93	174	0.58%	1.35%
F	123	123	129	77	0.66%	0.49%
G	129	113	167	49	0.81%	0.34%
H	130	125	272	74	1.32%	0.46%
K	110	105	22	22	0.13%	0.16%
Total	**1106**	**1090**	**1103**	**712**	**0.63%**	**0.51%**
Number of items (strokes)			158	128		

Therefore, for the main study the consortium recommended double entry procedures as part of a recruitment test for potential data entry operators, and as a means of monitoring the data entry personnel rather than as a compulsory procedure for data cleaning (Routitsky & Berezner, 2006).

Double coding of occupational data

Another new optional procedure for PISA 2006 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 2006 *Data Management Manua*[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 four-digit code had fewer discrepancies than those that required only a four-digit code. When analysing the double entry discrepancy reports from the English speaking countries the consortium found that when one of two coders entered both the description and code while another entered the code only, the discrepancy was mostly due to the second coder being incorrect. 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 (STQ-STF specific checks, ID checks questionnaires, ID checks occupation);
- Comparison of the identification variables in the test data (ID checks booklets, ID checks CBAS);
- Verification of the reliability data (reliability checks).

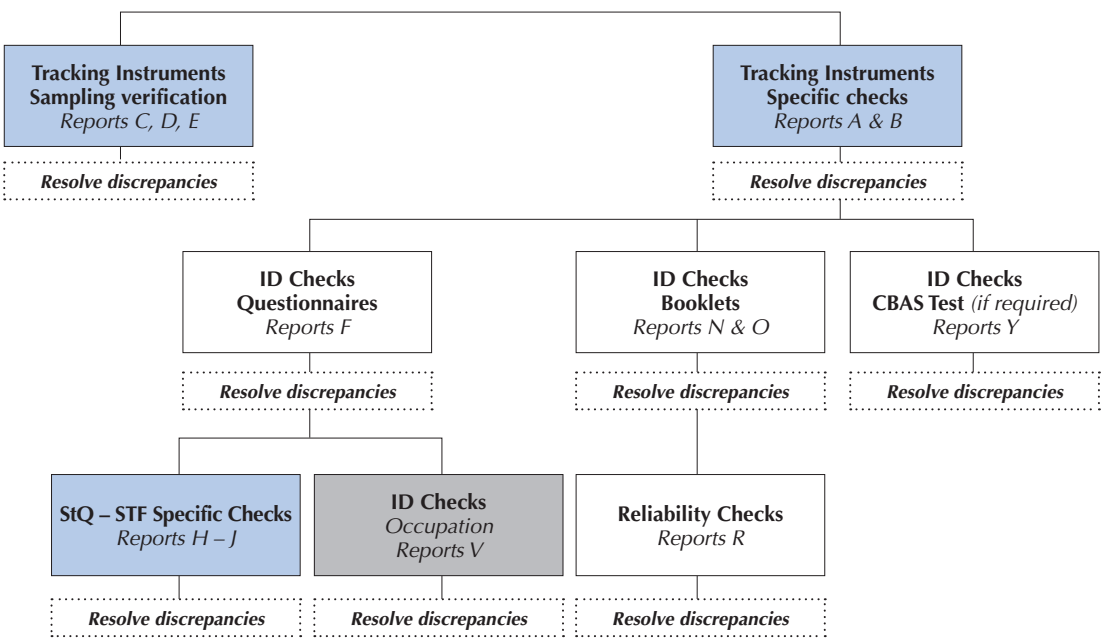
Some validity reports listed only incorrect records (*e.g.* students whose data were entered in more than one booklet instrument), whilst others listed both incorrect and suspicious records, which were records that could have been either correct or incorrect, but were deemed to be in need of confirmation.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

Figure 10.3

Validity reports – general hierarchy



PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

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 to find out whether this national variable require 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, the submitted data were recoded where necessary 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.

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 be 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. Students 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", (8400) 'Semi-skilled workers". Another example are additional auxiliary codes that were later recoded into 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.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Reader's Guide

Country codes – the following country codes are used in this report:

OECD countries

AUS	Australia
AUT	Austria
BEL	Belgium
BEF	Belgium (French Community)
BEN	Belgium (Flemish Community)
CAN	Canada
CAE	Canada (English Community)
CAF	Canada (French Community)
CZE	Czech Republic
DNK	Denmark
FIN	Finland
FRA	France
DEU	Germany
GRC	Greece
HUN	Hungary
ISL	Iceland
IRL	Ireland
ITA	Italy
JPN	Japan
KOR	Korea
LUX	Luxembourg
LXF	Luxembourg (French Community)
LXG	Luxembourg (German Community)
MEX	Mexico
NLD	Netherlands
NZL	New Zealand
NOR	Norway
POL	Poland
PRT	Portugal
SVK	Slovak Republic
ESP	Spain
ESB	Spain (Basque Community)
ESC	Spain (Catalonian Community)
ESS	Spain (Castillian Community)
SWE	Sweden
CHE	Switzerland
CHF	Switzerland (French Community)
CHG	Switzerland (German Community)
CHI	Switzerland (Italian Community)
TUR	Turkey
GBR	United Kingdom
IRL	Ireland
SCO	Scotland
USA	United States

Partner countries and economies

ARG	Argentina
AZE	Azerbaijan
BGR	Bulgaria
BRA	Brazil
CHL	Chile
COL	Colombia
EST	Estonia
HKG	Hong Kong-China
HRV	Croatia
IDN	Indonesia
JOR	Jordan
KGZ	Kyrgyztan
LIE	Liechtenstein
LTU	Lithuania
LVA	Latvia
LVL	Latvia (Latvian Community)
LVR	Latvia (Russian Community)
MAC	Macao-China
MNE	Montenegro
QAT	Qatar
ROU	Romania
RUS	Russian Federation
SRB	Serbia
SVN	Slovenia
TAP	Chinese Taipei
THA	Thailand
TUN	Tunisia
URY	Uruguay

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

References

Adams, R.J., Wilson, M. & **Wang, W.C.** (1997), The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement,* No. 21, pp. 1-23.

Adams, R.J., Wilson, M. R. & **Wu, M.L.** (1997), Multilevel item response models: An approach to errors in variables regression, *Journal of Educational and Behavioural Statistics,* No. 22 (1), pp. 46-75.

Adams, R.J. & **Wu, M.L.** (2002), *PISA 2000 Technical Report,* OECD, Paris.

Bollen, K.A. & **Long, S.J.** (1993) (eds.), *Testing Structural Equation Models*, Newbury Park: London.

Beaton, A.E. (1987), Implementing the new design: The NAEP 1983-84 technical report (Rep. No. 15-TR-20), Princeton, NJ: Educational Testing Service.

Buchmann, C. (2000), Family structure, parental perceptions and child labor in Kenya: What factors determine who is enrolled in school? *Soc. Forces,* No. 78, pp. 1349-79.

Buchmann, C. (2002), Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges, in Porter, A.C. and Gamoran, A. (eds.). *Methodological Advances in Cross-National Surveys of Educational Achievement* (pp. 150-97), Washington, DC: National Academy Press.

Creemers, B.P.M. (1994), *The Effective Classroom*, London: Cassell.

Cochran, W.G. (1977), *Sampling techniques,* third edition, New York, NY: John Wiley and Sons.

Ganzeboom, H.B.G., de Graaf, P.M. & **Treiman, D.J.** (1992), A standard international socio-economic index of occupational status, *Social Science Research*, No. 21, pp. 1-56.

Ganzeboom H.B. & **Treiman, D.J.** (1996), Internationally comparable measures of occupational status for the 1988 international standard classification of occupations, *Social Science Research*, No. 25, pp. 201-239.

Grisay, A. (2003), Translation procedures in OECD/PISA 2000 international assessment, *Language Testing,* No. 20 (2), pp. 225-240.

Hambleton, R.K., Swaminathan, H. & **Rogers, H.J.** (1991), *Fundamentals of item response theory*, Newbury Park, London, New Delhi: SAGE Publications.

Hambleton, R.K., Merenda, P.F. & **Spielberger, C.D.** (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment,* IEA Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Harkness, J.A., Van de Vijver, F.J.R. & **Mohler, P.Ph** (2003), *Cross-Cultural Survey Methods,* Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey.

Harvey-Beavis, A. (2002), Student and School Questionnaire Development, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report,* (pp. 33-38), OECD, Paris.

International Labour Organisation (ILO) (1990), *International Standard Classification of Occupations: ISCO-88.* Geneva: International Labour Office.

Jöreskog, K.G. & **Sörbom, Dag** (1993), *LISREL 8 User's Reference Guide*, Chicago: SSI.

Judkins, D.R. (1990), Fay's Method of Variance Estimation, *Journal of Official Statistics*, No. 6 (3), pp. 223-239.

Kaplan, D. (2000), *Structural equation modeling: Foundation and extensions,* Thousand Oaks: SAGE Publications.

Keyfitz, N. (1951), Sampling with probabilities proportionate to science: Adjustment for changes in probabilities, *Journal of the American Statistical Association,* No. 46, American Statistical Association, Alexandria, pp. 105-109.

Kish, L. (1992), Weighting for Unequal, *Pi. Journal of Official Statistics*, No. 8 (2), pp. 183-200.

LISREL (1993), K.G. Jöreskog & D. Sörbom, [computer software], Lincolnwood, IL: Scientific Software International, Inc.

Lohr, S.L. (1999), *Sampling: Design and Analysis*, Duxberry: Pacific Grove.

Macaskill, G., Adams, R.J. & **Wu, M.L.** (1998), Scaling methodology and procedures for the mathematics and science literacy, advanced mathematics and physics scale, in M. Martin and D.L. Kelly, Editors, *Third International Mathematics and Science Study, technical report Volume 3: Implementation and analysis*, Boston College, Chestnut Hill, MA.

Masters, G.N. & **Wright, B.D.** (1997), The Partial Credit Model, in W.J. van der Linden, & R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory* (pp. 101-122), New York/Berlin/Heidelberg: Springer.

Mislevy, R.J. (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika,* No. 56, pp. 177-196.

Mislevy, R.J., Beaton, A., Kaplan, B.A. & **Sheehan, K.** (1992), Estimating population characteristics from sparse matrix samples of item responses, *Journal of Educational Measurement,* No. 29 (2), pp. 133-161.

Mislevy, R.J. & **Sheehan, K.M.** (1987), Marginal estimation procedures, in Beaton, A.E., Editor, 1987. *The NAEP 1983-84 technical report,* National Assessment of Educational Progress, Educational Testing Service, Princeton, pp. 293-360.

Mislevy, R.J. & **Sheehan, K.M.** (1989), Information matrices in latent-variable models, *Journal of Educational Statistics,* No. 14, pp. 335-350.

Mislevy, R.J. & **Sheehan, K.M.** (1989), The role of collateral information about examinees in item parameter estimation, *Psychometrika,* No. 54, pp. 661-679.

Monseur, C. & **Berezner, A.** (2007), The Computation of Equating Errors in International Surveys in Education, *Journal of Applied Measurement,* No. 8 (3), 2007, pp. 323-335.

Monseur, C. (2005), An exploratory alternative approach for student non response weight adjustment, *Studies in Educational Evaluation,* No. 31 (2-3), pp. 129-144.

Muthen, B. & **L. Muthen** (1998), [computer software], *Mplus* Los Angeles, CA: Muthen & Muthen.

Muthen, B., du Toit, S.H.C. & **Spisic, D.** (1997), *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes,* unpublished manuscript.

OECD (1999), *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries,* OECD, Paris.

OECD (2003), *Literacy Skills for the World of Tomorrow: Further results from PISA 2000,* OECD, Paris.

OECD (2004), *Learning for Tomorrow's World – First Results from PISA 2003,* OECD, Paris.

OECD (2005), *Technical Report for the OECD Programme for International Student Assessment 2003,* OECD, Paris.

OECD (2006), *Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006,* OECD, Paris.

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World,* OECD, Paris.

PISA Consortium (2006), *PISA 2006 Main Study Data Management Manual, https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_data_management_manual.pdf*

Rasch, G. (1960), Probabilistic models for some intelligence and attainment tests, Copenhagen: Nielsen & Lydiche.

Routitski A. & **Berezner, A.** (2006), Issues influencing the validity of cross-national comparisons of student performance. Data Entry Quality and Parameter Estimation. Paper presented at the Annual Meeting of the American Educational Research Association (AERA) in San Francisco, 7-11 April, *https://mypisa.acer.edu.au/images/mypisadoc/aera06routitsky_berezner.pdf*

Rust, K. (1985), Variance Estimation for Complex Estimators in Sample Surveys, *Journal of Official Statistics,* No. 1, pp. 381-397.

Rust, K.F. & **Rao, J.N.K.** (1996), Variance Estimation for Complex Surveys Using Replication Techniques, *Survey Methods in Medical Research,* No. 5, pp. 283-310.

Shao, J. (1996), Resampling Methods in Sample Surveys (with Discussion), *Statistics,* No. 27, pp. 203-254.

Särndal, C.-E., Swensson, B. & **Wretman, J.** (1992), *Model Assisted Survey Sampling,* New York: Springer-Verlag.

SAS® CALIS (1992), W. Hartmann [computer software], Cary, NC: SAS Institute Inc.

Scheerens, J. (1990), School effectiveness and the development of process indicators of school functioning, *School effectiveness and school improvement,* No. 1, pp. 61-80.

Scheerens, J. & **Bosker, R.J.** (1997), *The Foundations of School Effectiveness,* Oxford: Pergamon.

Schulz, W. (2002), Constructing and Validating the Questionnaire composites, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report,* OECD, Paris.

Schulz, W. (2004), Mapping Student Scores to Item Responses, in W. Schulz and H. Sibberns (eds.), *IEA Civic Education Study, Technical Report* (pp. 127-132), Amsterdam: IEA.

Schulz, W. (2006a), *Testing Parameter Invariance for Questionnaire Indices using Confirmatory Factor Analysis and Item Response Theory,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Schulz, W. (2006b), *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Thorndike, R.L. (1973), *Reading comprehension in fifteen countries,* New York, Wiley: and Stockholm: Almqvist & Wiksell.

Travers, K.J. & **Westbury, I.** (1989), *The IEA Study of Mathematics I: Analysis of Mathematics Curricula,* Oxford: Pergamon Press.

Travers, K.J., Garden R.A. & **Rosier, M.** (1989), Introduction to the Study, in Robitaille, D. A. and Garden, R. A. (eds), *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics Curricula,* Oxford: Pergamon Press.

Verhelst, N. (2002), Coder and Marker Reliabilaity Studies, in R.J. Adams & M.L. Wu (eds.), *PISA 2000 Technical Report*. OECD, Paris.

Walberg, H.J. (1984), Improving the productivity of American schools, *Educational Leadership,* No. 41, pp. 19-27.

Walberg, H. (1986), Synthesis of research on teaching, in M. Wittrock (ed.), *Handbook of research on teaching* (pp. 214-229), New York: Macmillan.

Walker, M. (2006), *The choice of Likert or dichotomous items to measure attitudes across culturally distinct countries in international comparative educational research*. Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Walker, M. (2007), Ameliorating Culturally-Based Extreme Response Tendencies To Attitude items, *Journal of Applied Measurement,* No. 8, pp. 267-278.

Warm, T.A. (1989), Weighted Likelihood Estimation of Ability in Item Response Theory, *Psychometrika*, No. 54 (3), pp. 427-450.

Westat (2007), *WesVar® 5.1* Computer software and manual, Rockville, MD: Author (also see *http://www.westat.com/wesvar/*).

Wilson, M. (1994), Comparing Attitude Across Different Cultures: Two Quantitative Approaches to Construct Validity, in M. Wilson (ed.), *Objective measurement II: Theory into practice* (pp. 271-292), Norwood, NJ: Ablex.

Wolter, K.M. (2007), *Introduction to Variance Estimation*. Second edition, Springer: New York.

Wu, M.L., Adams, R.J. & **Wilson, M.R.** (1997), *ConQuest®: Multi-Aspect Test Software* [computer program manual], Camberwell, Vic.: Australian Council for Educational Research.

List of abbreviations – the following abbreviations are used in this report:

ACER	Australian Council for Educational Research
AGFI	Adjusted Goodness-of-Fit Index
BRR	Balanced Repeated Replication
CBAS	Computer Based Assessment of Science
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CITO	National Institute for Educational Measurement, The Netherlands
CIVED	Civic Education Study
DIF	Differential Item Functioning
ENR	Enrolment of 15-year-olds
ESCS	PISA Index of Economic, Social and Cultural Status
ETS	Educational Testing Service
IAEP	International Assessment of Educational Progress
I	Sampling Interval
ICR	Inter-Country Coder Reliability Study
ICT	Information Communication Technology
IEA	International Association for the Evaluation of Educational Achievement
INES	OECD Indicators of Education Systems
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISEI	International Socio-Economic Index
MENR	Enrolment for moderately small school
MOS	Measure of size
NCQM	National Centre Quality Monitor
NDP	National Desired Population
NEP	National Enrolled Population
NFI	Normed Fit Index
NIER	National Institute for Educational Research, Japan
NNFI	Non-Normed Fit Index
NPM	National Project Manager
OECD	Organisation for Economic Cooperation and Development
PISA	Programme for International Student Assessment
PPS	Probability Proportional to Size
PGB	PISA Governing Board
PQM	PISA Quality Monitor
PSU	Primary Sampling Units
QAS	Questionnaire Adaptations Spreadsheet
RMSEA	Root Mean Square Error of Approximation
RN	Random Number
SC	School Co-ordinator
SE	Standard Error
SD	Standard Deviation
SEM	Structural Equation Modelling
SMEG	Subject Matter Expert Group
SPT	Study Programme Table
TA	Test Administrator
TAG	Technical Advisory Group
TCS	Target Cluster Size
TIMSS	Third International Mathematics and Science Study
TIMSS-R	Third International Mathematics and Science Study – Repeat
VENR	Enrolment for very small schools
WLE	Weighted Likelihood Estimates

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Table of contents

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

LIST OF BOXES

LIST OF FIGURES

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

LIST OF TABLES

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009