

Data Management Procedures

Introduction.....	164
<i>KeyQuest</i>	167
Data management at the national centre.....	167
▪ National modifications to the database.....	167
▪ Student sampling with <i>KeyQuest</i>	167
▪ Data entry quality control.....	167
Data cleaning at ACER.....	171
▪ Recoding of national adaptations.....	171
▪ Data cleaning organisation.....	171
▪ Cleaning reports.....	171
▪ General recodings.....	171
Final review of the data.....	172
▪ Review of the test and questionnaire data.....	172
▪ Review of the sampling data.....	172
Next steps in preparing the international database.....	172



INTRODUCTION

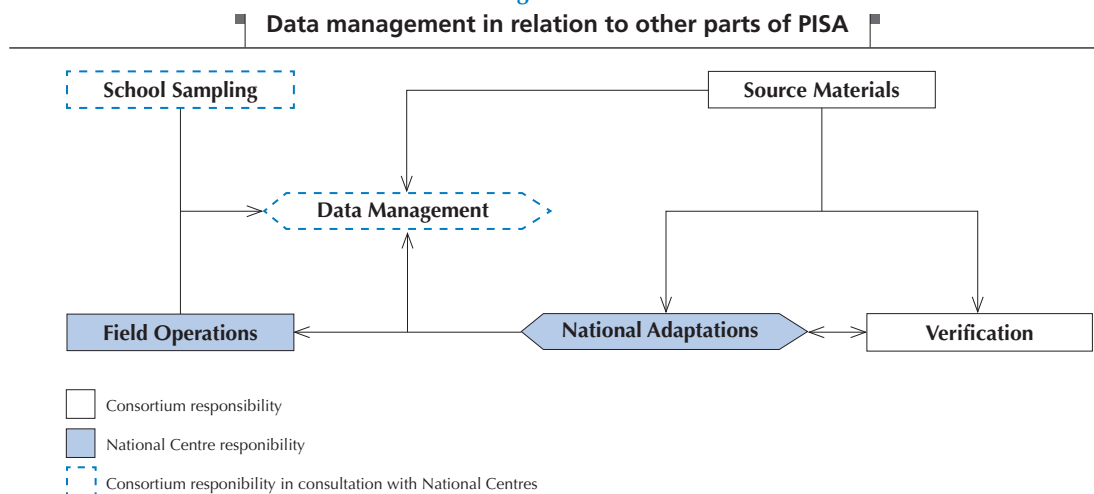
The PISA assessment establishes certain data collection requirements that are common to all PISA participants. Test instruments include the same test items in all participating countries, and data collection procedures are applied in a common and consistent way amongst all participants to help ensure 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 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.

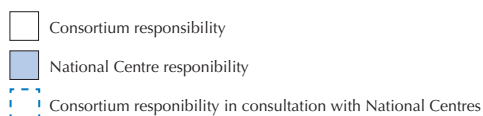
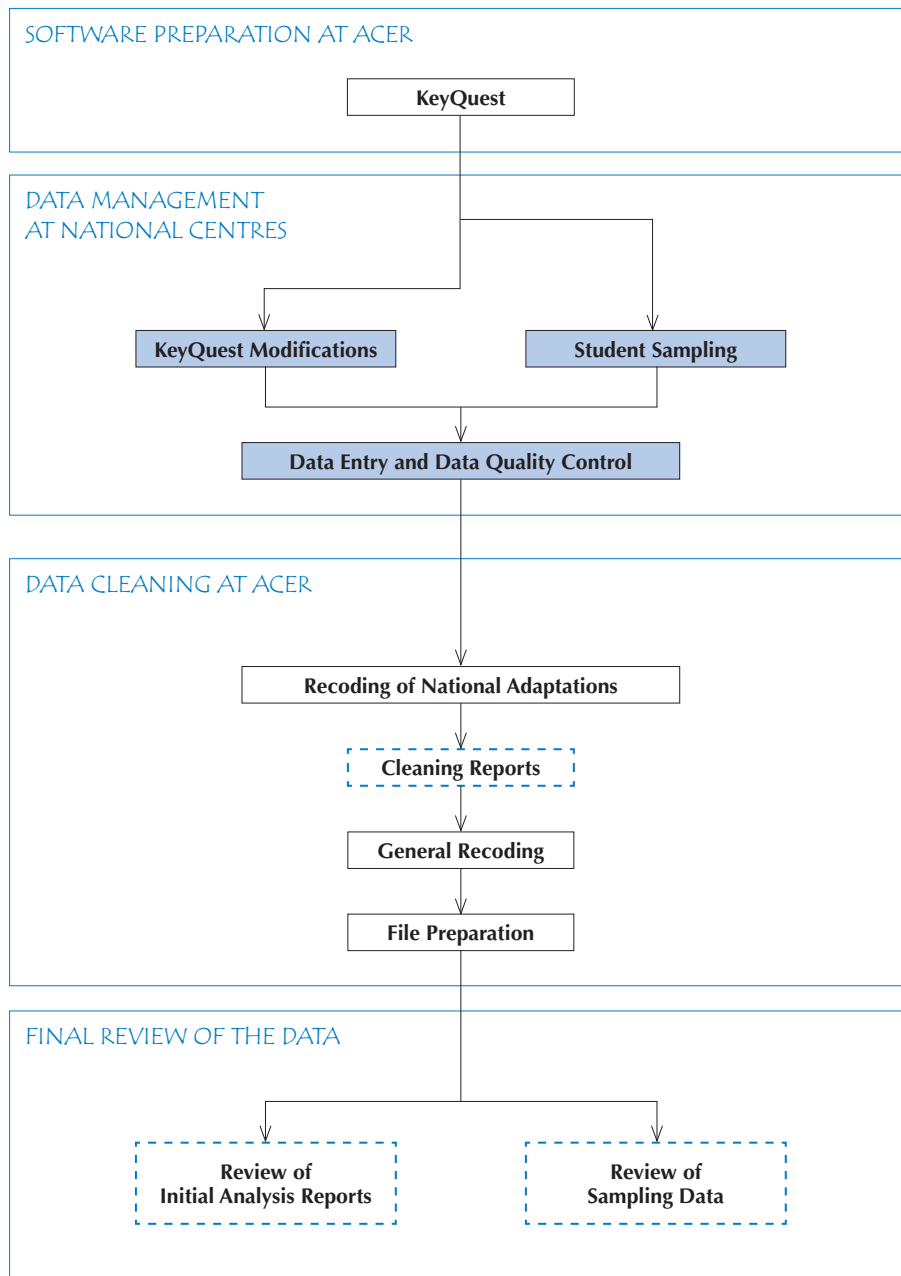
Once ACER has received all cleaning reports from the national centres and has introduced into the database all corrections recommended in these reports, a number of general rules are applied to the small number of unresolved inconsistencies in the PISA database.

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.



Figure 10.2

Major data management stages in PISA





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.



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

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

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

Country	Number of students		Number of discrepancies		D	
	Booklet					
	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 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 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).

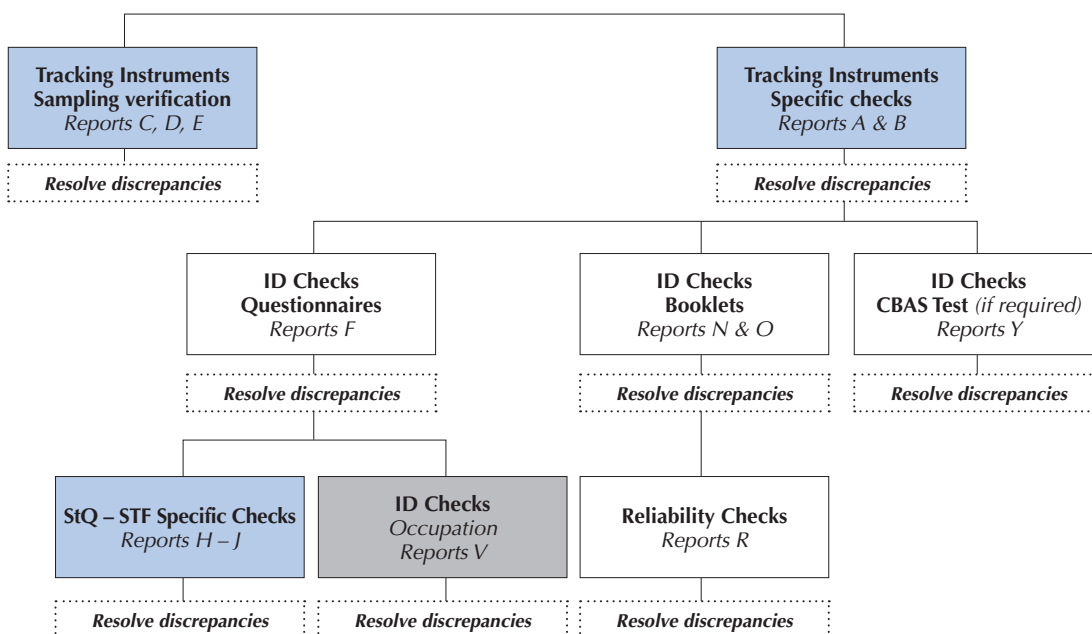
Some validity reports listed only incorrect records (e.g. students whose data were entered in more than one booklet instrument), whilst others listed both incorrect and suspicious records, which were records that could have been either correct or incorrect, but were deemed to be in need of confirmation.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





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.



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	Kyrgyzstan
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



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/era06routitsky_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 Reliability 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	NPM	National Project Manager
AGFI	Adjusted Goodness-of-Fit Index	OECD	Organisation for Economic Cooperation and Development
BRR	Balanced Repeated Replication	PISA	Programme for International Student Assessment
CBAS	Computer Based Assessment of Science	PPS	Probability Proportional to Size
CFA	Confirmatory Factor Analysis	PGB	PISA Governing Board
CFI	Comparative Fit Index	PQM	PISA Quality Monitor
CITO	National Institute for Educational Measurement, The Netherlands	PSU	Primary Sampling Units
CIVED	Civic Education Study	QAS	Questionnaire Adaptations Spreadsheet
DIF	Differential Item Functioning	RMSEA	Root Mean Square Error of Approximation
ENR	Enrolment of 15-year-olds	RN	Random Number
ESCS	PISA Index of Economic, Social and Cultural Status	SC	School Co-ordinator
ETS	Educational Testing Service	SE	Standard Error
IAEP	International Assessment of Educational Progress	SD	Standard Deviation
I	Sampling Interval	SEM	Structural Equation Modelling
ICR	Inter-Country Coder Reliability Study	SMEG	Subject Matter Expert Group
ICT	Information Communication Technology	SPT	Study Programme Table
IEA	International Association for the Evaluation of Educational Achievement	TA	Test Administrator
INES	OECD Indicators of Education Systems	TAG	Technical Advisory Group
IRT	Item Response Theory	TCS	Target Cluster Size
ISCED	International Standard Classification of Education	TIMSS	Third International Mathematics and Science Study
ISCO	International Standard Classification of Occupations	TIMSS-R	Third International Mathematics and Science Study – Repeat
ISEI	International Socio-Economic Index	VENR	Enrolment for very small schools
MENR	Enrolment for moderately small school	WLE	Weighted Likelihood Estimates
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		



Table of contents

FOREWORD	3
CHAPTER 1 PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT: AN OVERVIEW	19
Participation	21
Features of PISA	22
Managing and implementing PISA	23
Organisation of this report	23
READER'S GUIDE	25
CHAPTER 2 TEST DESIGN AND TEST DEVELOPMENT	27
Test scope and format	28
Test design	28
Test development centres	29
Development timeline	30
The PISA 2006 scientific literacy framework	30
Test development – cognitive items	31
▪ Item development process.....	31
▪ National item submissions.....	33
▪ National review of items.....	34
▪ International item review.....	35
▪ Preparation of dual (English and French) source versions.....	35
Test development – attitudinal items	35
Field trial	38
▪ Field trial selection.....	38
▪ Field trial design.....	39
▪ Despatch of field trial instruments.....	40
▪ Field trial coder training.....	40
▪ Field trial coder queries.....	40
▪ Field trial outcomes.....	41
▪ National review of field trial items.....	42
Main study	42
▪ Main study science items.....	42
▪ Main study reading items.....	44
▪ Main study mathematics items.....	45
▪ Despatch of main study instruments.....	46
▪ Main study coder training.....	46
▪ Main study coder query service.....	46
▪ Review of main study item analyses.....	47



CHAPTER 3 THE DEVELOPMENT OF THE PISA CONTEXT QUESTIONNAIRES	49
Overview	50
The conceptual structure	51
▪ A conceptual framework for PISA 2006	51
Research areas in PISA 2006	55
The development of the context questionnaires	57
The coverage of the questionnaire material	58
▪ Student questionnaire	58
▪ School questionnaire	59
▪ International options	59
▪ National questionnaire material	60
The implementation of the context questionnaires	60
 CHAPTER 4 SAMPLE DESIGN	 63
Target population and overview of the sampling design	64
Population coverage, and school and student participation rate standards	65
▪ Coverage of the PISA international target population	65
▪ Accuracy and precision	66
▪ School response rates	66
▪ Student response rates	68
Main study school sample	68
▪ Definition of the national target population	68
▪ The sampling frame	69
▪ Stratification	70
▪ Assigning a measure of size to each school	74
▪ School sample selection	74
▪ PISA and TIMSS or PIRLS overlap control	76
▪ Student samples	82
 CHAPTER 5 TRANSLATION AND CULTURAL APPROPRIATENESS OF THE TEST AND SURVEY MATERIAL	 85
Introduction	86
Development of source versions	86
Double translation from two source languages	87
PISA translation and adaptation guidelines	88
Translation training session	89
Testing languages and translation/adaptation procedures	89
International verification of the national versions	91
▪ VegaSuite	93
▪ Documentation	93
▪ Verification of test units	93
▪ Verification of the booklet shell	94
▪ Final optical check	94
▪ Verification of questionnaires and manuals	94
▪ Final check of coding guides	95
▪ Verification outcomes	95



Translation and verification outcomes – national version quality	96
▪ Analyses at the country level.....	96
▪ Analyses at the item level.....	103
▪ Summary of items lost at the national level, due to translation, printing or layout errors.....	104
CHAPTER 6 FIELD OPERATIONS	105
Overview of roles and responsibilities	106
▪ National project managers.....	106
▪ School coordinators.....	107
▪ Test administrators.....	107
▪ School associates.....	108
The selection of the school sample	108
Preparation of test booklets, questionnaires and manuals	108
The selection of the student sample	109
Packaging and shipping materials	110
Receipt of materials at the national centre after testing	110
Coding of the tests and questionnaires	111
▪ Preparing for coding.....	111
▪ Logistics prior to coding.....	113
▪ Single coding design.....	115
▪ Multiple coding.....	117
▪ Managing the process coding.....	118
▪ Cross-national coding.....	120
▪ Questionnaire coding.....	120
Data entry, data checking and file submission	120
▪ Data entry.....	120
▪ Data checking.....	120
▪ Data submission.....	121
▪ After data were submitted.....	121
The main study review	121
CHAPTER 7 QUALITY ASSURANCE	123
PISA quality control	124
▪ Comprehensive operational manuals.....	124
▪ National level implementation planning document.....	124
PISA quality monitoring	124
▪ Field trial and main study review.....	124
▪ Final optical check.....	126
▪ National centre quality monitor (NCQM) visits.....	126
▪ PISA quality monitor (PQM) visits.....	126
▪ Test administration.....	127
▪ Delivery.....	128
CHAPTER 8 SURVEY WEIGHTING AND THE CALCULATION OF SAMPLING VARIANCE	129
Survey weighting	130
The school base weight	131
▪ The school weight trimming factor.....	132

<ul style="list-style-type: none"> ▪ The student base weight 132 ▪ School non-response adjustment..... 132 ▪ Grade non-response adjustment..... 134 ▪ Student non-response adjustment..... 135 ▪ Trimming student weights..... 136 ▪ Comparing the PISA 2006 student non-response adjustment strategy with the strategy used for PISA 2003 136 ▪ The comparison..... 138 	
Calculating sampling variance	139
<ul style="list-style-type: none"> ▪ The balanced repeated replication variance estimator..... 139 ▪ Reflecting weighting adjustments..... 141 ▪ Formation of variance strata..... 141 ▪ Countries where all students were selected for PISA..... 141 	
CHAPTER 9 SCALING PISA COGNITIVE DATA	143
The mixed coefficients multinomial logit model	144
<ul style="list-style-type: none"> ▪ The population model..... 145 ▪ Combined model..... 146 	
Application to PISA	146
<ul style="list-style-type: none"> ▪ National calibrations..... 146 ▪ National reports..... 147 ▪ International calibration 153 ▪ Student score generation..... 153 	
Booklet effects	155
Analysis of data with plausible values	156
Developing common scales for the purposes of trends	157
<ul style="list-style-type: none"> ▪ Linking PISA 2003 and PISA 2006 for reading and mathematics 158 ▪ Uncertainty in the link..... 158 	
CHAPTER 10 DATA MANAGEMENT PROCEDURES	163
Introduction	164
KeyQuest	167
Data management at the national centre	167
<ul style="list-style-type: none"> ▪ National modifications to the database 167 ▪ Student sampling with <i>KeyQuest</i>..... 167 ▪ Data entry quality control 167 	
Data cleaning at ACER	171
<ul style="list-style-type: none"> ▪ Recoding of national adaptations..... 171 ▪ Data cleaning organisation..... 171 ▪ Cleaning reports..... 171 ▪ General recodings..... 171 	
Final review of the data	172
<ul style="list-style-type: none"> ▪ Review of the test and questionnaire data 172 ▪ Review of the sampling data 172 	
Next steps in preparing the international database	172



CHAPTER 11 SAMPLING OUTCOMES	175
Design effects and effective sample sizes	187
▪ Variability of the design effect.....	191
▪ Design effects in PISA for performance variables.....	191
Summary analyses of the design effect	203
▪ Countries with outlying standard errors.....	205
 CHAPTER 12 SCALING OUTCOMES	 207
International characteristics of the item pool	208
▪ Test targeting.....	208
▪ Test reliability.....	208
▪ Domain inter-correlations.....	208
▪ Science scales.....	215
Scaling outcomes	216
▪ National item deletions.....	216
▪ International scaling.....	219
▪ Generating student scale scores.....	219
Test length analysis	219
Booklet effects	221
▪ Overview of the PISA cognitive reporting scales.....	232
▪ PISA overall literacy scales.....	234
▪ PISA literacy scales.....	234
▪ Special purpose scales.....	234
Observations concerning the construction of the PISA overall literacy scales	235
▪ Framework development.....	235
▪ Testing time and item characteristics.....	236
▪ Characteristics of each of the links.....	237
Transforming the plausible values to PISA scales	246
▪ Reading.....	246
▪ Mathematics.....	246
▪ Science.....	246
▪ Attitudinal scales.....	247
Link error	247
 CHAPTER 13 CODING AND MARKER RELIABILITY STUDIES	 249
Homogeneity analyses	251
Multiple marking study outcomes (variance components)	254
▪ Generalisability coefficients.....	254
International coding review	261
▪ Background to changed procedures for PISA 2006.....	261
▪ ICR procedures.....	261
▪ Outcomes.....	264
▪ Cautions.....	270



CHAPTER 14 DATA ADJUDICATION	271
Introduction	272
▪ Implementing the standards – quality assurance	272
▪ Information available for adjudication	273
▪ Data adjudication process	273
General outcomes	274
▪ Overview of response rate issues	274
▪ Detailed country comments	275
 CHAPTER 15 PROFICIENCY SCALE CONSTRUCTION	 283
Introduction	284
Development of the described scales	285
▪ Stage 1: Identifying possible scales	285
▪ Stage 2: Assigning items to scales	286
▪ Stage 3: Skills audit	286
▪ Stage 4: Analysing field trial data	286
▪ Stage 5: Defining the dimensions	287
▪ Stage 6: Revising and refining with main study data	287
▪ Stage 7: Validating	287
Defining proficiency levels	287
Reporting the results for PISA science	290
▪ Building an item map	290
▪ Levels of scientific literacy	292
▪ Interpreting the scientific literacy levels	299
 CHAPTER 16 SCALING PROCEDURES AND CONSTRUCT VALIDATION OF CONTEXT QUESTIONNAIRE DATA	 303
Overview	304
Simple questionnaire indices	304
▪ Student questionnaire indices	304
▪ School questionnaire indices	307
▪ Parent questionnaire indices	309
Scaling methodology and construct validation	310
▪ Scaling procedures	310
▪ Construct validation	312
▪ Describing questionnaire scale indices	314
Questionnaire scale indices	315
▪ Student scale indices	315
▪ School questionnaire scale indices	340
▪ Parent questionnaire scale indices	342
▪ The PISA index of economic, social and cultural status (ESCS)	346
 CHAPTER 17 VALIDATION OF THE EMBEDDED ATTITUDINAL SCALES	 351
Introduction	352
International scalability	353
▪ Analysis of item dimensionality with exploratory and confirmatory factor analysis	353
▪ Fit to item response model	353



▪ Reliability.....	355
▪ Differential item functioning.....	355
▪ Summary of scalability.....	357
Relationship and comparisons with other variables.....	357
▪ Within-country student level correlations with achievement and selected background variables.....	358
▪ Relationships between embedded scales and questionnaire.....	360
▪ Country level correlations with achievement and selected background variables.....	361
▪ Variance decomposition.....	363
▪ Observations from other cross-national data collections.....	363
▪ Summary of relations with other variables.....	364
Conclusion.....	364
CHAPTER 18 INTERNATIONAL DATABASE.....	367
Files in the database.....	368
▪ Student files.....	368
▪ School file.....	370
▪ Parent file.....	370
Records in the database.....	371
▪ Records included in the database.....	371
▪ Records excluded from the database.....	371
Representing missing data.....	371
How are students and schools identified?.....	372
Further information.....	373
REFERENCES.....	375
APPENDICES.....	379
Appendix 1 PISA 2006 main study item pool characteristics.....	380
Appendix 2 Contrast coding used in conditioning.....	389
Appendix 3 Design effect tables.....	399
Appendix 4 Changes to core questionnaire items from 2003 to 2006.....	405
Appendix 5 Mapping of ISCED to years.....	411
Appendix 6 National household possession items.....	412
Appendix 7 Exploratory and confirmatory factor analyses for the embedded items.....	414
Appendix 8 PISA consortium, staff and consultants.....	416



LIST OF BOXES

Box 1.1	Core features of PISA 2006.....	22
---------	---------------------------------	----

LIST OF FIGURES

Figure 2.1	Main study Interest in Science item.....	36
Figure 2.2	Main study Support for Scientific Enquiry item.....	36
Figure 2.3	Field trial Match-the-opinion Responsibility item.....	37
Figure 3.1	Conceptual grid of variable types.....	52
Figure 3.2	The two-dimensional conceptual matrix with examples of variables collected or available from other sources.....	54
Figure 4.1	School response rate standard.....	67
Figure 6.1	Design for the single coding of science and mathematics.....	115
Figure 6.2	Design for the single coding of reading.....	116
Figure 9.1	Example of item statistics in Report 1.....	148
Figure 9.2	Example of item statistics in Report 2.....	149
Figure 9.3	Example of item statistics shown in Graph B.....	150
Figure 9.4	Example of item statistics shown in Graph C.....	151
Figure 9.5	Example of item statistics shown in Table D.....	151
Figure 9.6	Example of summary of dodgy items for a country in Report 3a.....	152
Figure 9.7	Example of summary of dodgy items in Report 3b.....	152
Figure 10.1	Data management in relation to other parts of PISA.....	164
Figure 10.2	Major data management stages in PISA.....	166
Figure 10.3	Validity reports - general hierarchy.....	170
Figure 11.1	Standard error on a mean estimate depending on the intraclass correlation.....	188
Figure 11.2	Relationship between the standard error for the science performance mean and the intraclass correlation within explicit strata (PISA 2006).....	205
Figure 12.1	Item plot for mathematics items.....	210
Figure 12.2	Item plot for reading items.....	211
Figure 12.3	Item plot for science items.....	212
Figure 12.4	Item plot for interest items.....	213
Figure 12.5	Item plot for support items.....	214
Figure 12.6	Scatter plot of per cent correct for reading link items in PISA 2000 and PISA 2003.....	238
Figure 12.7	Scatter plot of per cent correct for reading link items in PISA 2003 and PISA 2006.....	240
Figure 12.8	Scatter plot of per cent correct for mathematics link items in PISA 2003 and PISA 2006.....	242
Figure 12.9	Scatter plot of per cent correct for science link items in PISA 2000 and PISA 2003.....	244
Figure 12.10	Scatter plot of per cent correct for science link items in PISA 2003 and PISA 2006.....	245



Figure 13.1	Variability of the homogeneity indices for science items in field trial	250
Figure 13.2	Average of the homogeneity indices for science items in field trial and main study	251
Figure 13.3	Variability of the homogeneity indices for each science item in the main study	252
Figure 13.4	Variability of the homogeneity indices for each reading item in the main study	252
Figure 13.5	Variability of the homogeneity indices for each mathematics item	252
Figure 13.6	Variability of the homogeneity indices for the participating countries in the main study	253
Figure 13.7	Example of ICR report (reading)	269
<hr/>		
Figure 14.1	Attained school response rates	274
<hr/>		
Figure 15.1	The relationship between items and students on a proficiency scale	285
Figure 15.2	What it means to be at a level	289
Figure 15.3	A map for selected science items	291
Figure 15.4	Summary descriptions of the six proficiency levels on the science scale	294
Figure 15.5	Summary descriptions of six proficiency levels in <i>identifying scientific issues</i>	295
Figure 15.6	Summary descriptions of six proficiency levels in <i>explaining phenomena scientifically</i>	297
Figure 15.7	Summary descriptions of six proficiency levels in <i>using scientific evidence</i>	300
<hr/>		
Figure 16.1	Summed category probabilities for fictitious item	314
Figure 16.2	Fictitious example of an item map	315
Figure 16.3	Scatterplot of country means for ESCS 2003 and ESCS 2006	347
<hr/>		
Figure 17.1	Distribution of item fit mean square statistics for embedded attitude items	354
Figure 17.2	An example of the ESC plot for item S408RNA	356
Figure 17.3	Scatterplot of mean mathematics interest against mean mathematics for PISA 2003	363

LIST OF TABLES

Table 1.1	PISA 2006 participants	21
<hr/>		
Table 2.1	Cluster rotation design used to form test booklets for PISA 2006	29
Table 2.2	Test development timeline for PISA 2006	30
Table 2.3	Science field trial all items	39
Table 2.4	Allocation of item clusters to test booklets for field trial	39
Table 2.5	Science main study items (item format by competency)	43
Table 2.6	Science main study items (item format by knowledge type)	44
Table 2.7	Science main study items (knowledge category by competency)	44
Table 2.8	Reading main study items (item format by aspect)	44
Table 2.9	Reading main study items (item format by text format)	45
Table 2.10	Reading main study items (text type by aspect)	45
Table 2.11	Mathematics main study items (item format by competency cluster)	45
Table 2.12	Mathematics main study items (item format by content category)	46
Table 2.13	Mathematics main study items (content category by competency cluster)	46

Table 3.1	Themes and constructs/variables in PISA 2006.....	56
Table 4.1	Stratification variables	71
Table 4.2	Schedule of school sampling activities	78
Table 5.1	Countries sharing a common version with national adaptations	90
Table 5.2	PISA 2006 translation/adaptation procedures.....	91
Table 5.3	Mean deviation and root mean squared error of the item by country interactions for each version.....	97
Table 5.4	Correlation between national item parameter estimates for Arabic versions.....	99
Table 5.5	Correlation between national item parameter estimates for Chinese versions.....	99
Table 5.6	Correlation between national item parameter estimates for Dutch versions.....	99
Table 5.7	Correlation between national item parameter estimates for English versions.....	99
Table 5.8	Correlation between national item parameter estimates for French versions.....	99
Table 5.9	Correlation between national item parameter estimates for German versions.....	100
Table 5.10	Correlation between national item parameter estimates for Hungarian versions.....	100
Table 5.11	Correlation between national item parameter estimates for Italian versions.....	100
Table 5.12	Correlation between national item parameter estimates for Portuguese versions.....	100
Table 5.13	Correlation between national item parameter estimates for Russian versions.....	100
Table 5.14	Correlation between national item parameter estimates for Spanish versions	100
Table 5.15	Correlation between national item parameter estimates for Swedish versions	100
Table 5.16	Correlation between national item parameter estimates within countries.....	101
Table 5.17	Variance estimate.....	102
Table 5.18	Variance estimates	103
Table 6.1	Design for the multiple coding of science and mathematics.....	118
Table 6.2	Design for the multiple coding of reading.....	118
Table 8.1	Non-response classes	133
Table 9.1	Deviation contrast coding scheme	154
Table 10.1	Double entry discrepancies per country: field trial data.....	169
Table 11.1	Sampling and coverage rates.....	178
Table 11.2	School response rates before replacement.....	182
Table 11.3	School response rates after replacement.....	184
Table 11.4	Student response rates after replacement.....	185
Table 11.5	Standard errors for the PISA 2006 combined science scale	189
Table 11.6	Design effect 1 by country, by domain and cycle.....	193
Table 11.7	Effective sample size 1 by country, by domain and cycle.....	194
Table 11.8	Design effect 2 by country, by domain and cycle.....	195
Table 11.9	Effective sample size 2 by country, by domain and cycle.....	196
Table 11.10	Design effect 3 by country, by domain and by cycle.....	197



Table 11.11	Effective sample size 3 by country, by domain and cycle	198
Table 11.12	Design effect 4 by country, by domain and cycle.....	199
Table 11.13	Effective sample size 4 by country, by domain and cycle	200
Table 11.14	Design effect 5 by country, by domain and cycle.....	201
Table 11.15	Effective sample size 5 by country, by domain and cycle	202
Table 11.16	Median of the design effect 3 per cycle and per domain across the 35 countries that participated in every cycle.....	203
Table 11.17	Median of the standard errors of the student performance mean estimate for each domain and PISA cycle for the 35 countries that participated in every cycle	203
Table 11.18	Median of the number of participating schools for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.19	Median of the school variance estimate for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.20	Median of the intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.21	Median of the within explicit strata intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle	205
Table 11.22	Median of the percentages of school variances explained by explicit stratification variables, for each domain and PISA cycle for the 35 countries that participated in every cycle	205
<hr/>		
Table 12.1	Number of sampled student by country and booklet.....	209
Table 12.2	Reliabilities of each of the four overall scales when scaled separately.....	215
Table 12.3	Latent correlation between the five domains	215
Table 12.4	Latent correlation between science scales	215
Table 12.5	Items deleted at the national level	216
Table 12.6	Final reliability of the PISA scales	216
Table 12.7	National reliabilities for the main domains.....	217
Table 12.8	National reliabilities for the science subscales.....	218
Table 12.9	Average number of not-reached items and missing items by booklet.....	219
Table 12.10	Average number of not-reached items and missing items by country.....	220
Table 12.11	Distribution of not-reached items by booklet	221
Table 12.12	Estimated booklet effects on the PISA scale.....	221
Table 12.13	Estimated booklet effects in logits	221
Table 12.14	Variance in mathematics booklet means	222
Table 12.15	Variance in reading booklet means.....	224
Table 12.16	Variance in science booklet means.....	226
Table 12.17	Variance in interest booklet means	228
Table 12.18	Variance in support booklet means.....	230
Table 12.19	Summary of PISA cognitive reporting scales	233
Table 12.20	Linkage types among PISA domains 2000-2006	235
Table 12.21	Number of unique item minutes for each domain for each PISA assessments.....	237
Table 12.22	Numbers of link items between successive PISA assessments.....	237
Table 12.23	Per cent correct for reading link items in PISA 2000 and PISA 2003	238
Table 12.24	Per cent correct for reading link items in PISA 2003 and PISA 2006	239
Table 12.25	Per cent correct for mathematics link items in PISA 2003 and PISA 2006	241

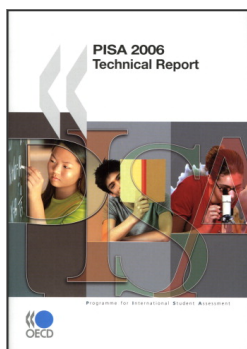
Table 12.26	Per cent correct for science link items in PISA 2000 and PISA 2003	243
Table 12.27	Per cent correct for science link items in PISA 2003 and PISA 2006	245
Table 12.28	Link error estimates	247
<hr/>		
Table 13.1	Variance components for mathematics.....	255
Table 13.2	Variance components for science	256
Table 13.3	Variance components for reading.....	257
Table 13.4	Generalisability estimates for mathematics.....	258
Table 13.5	Generalisability estimates for science	259
Table 13.6	Generalisability estimates for reading	260
Table 13.7	Examples of flagged cases	263
Table 13.8	Count of analysis groups showing potential bias, by domain.....	264
Table 13.9	Comparison of codes assigned by verifier and adjudicator	265
Table 13.10	Outcomes of ICR analysis part 1	265
Table 13.11	ICR outcomes by country and domain	266
<hr/>		
Table 15.1	Scientific literacy performance band definitions on the PISA scale	293
<hr/>		
Table 16.1	ISCO major group white-collar/blue-collar classification	306
Table 16.2	ISCO occupation categories classified as science-related occupations	307
Table 16.3	OECD means and standard deviations of WL estimates	311
Table 16.4	Median, minimum and maximum percentages of between-school variance for student-level indices across countries.....	313
Table 16.5	Household possessions and home background indices.....	316
Table 16.6	Scale reliabilities for home possession indices in OECD countries	317
Table 16.7	Scale reliabilities for home possession indices in partner countries/economies	318
Table 16.8	Item parameters for interest in science learning (INTSCIE).....	318
Table 16.9	Item parameters for enjoyment of science (JOYSCIE)	319
Table 16.10	Model fit and estimated latent correlations for interest in and enjoyment of science learning.....	319
Table 16.11	Scale reliabilities for interest in and enjoyment of science learning.....	320
Table 16.12	Item parameters for instrumental motivation to learn science (INSTSCIE).....	320
Table 16.13	Item parameters for future-oriented science motivation (SCIEFUT).....	321
Table 16.14	Model fit and estimated latent correlations for motivation to learn science	321
Table 16.15	Scale reliabilities for instrumental and future-oriented science motivation.....	322
Table 16.16	Item parameters for science self-efficacy (SCIEEFF).....	322
Table 16.17	Item parameters for science self-concept (SCSCIE).....	323
Table 16.18	Model fit and estimated latent correlations for science self-efficacy and science self-concept.....	323
Table 16.19	Scale reliabilities for science self-efficacy and science self-concept.....	324
Table 16.20	Item parameters for general value of science (GENSCIE).....	324
Table 16.21	Item parameters for personal value of science (PERSCIE).....	325
Table 16.22	Model fit and estimated latent correlations for general and personal value of science.....	325
Table 16.23	Scale reliabilities for general and personal value of science.....	326
Table 16.24	Item parameters for science activities (SCIEACT)	326



Table 16.25	Scale reliabilities for the science activities index	327
Table 16.26	Item parameters for awareness of environmental issues (ENVAWARE)	327
Table 16.27	Item parameters for perception of environmental issues (ENVPERC)	328
Table 16.28	Item parameters for environmental optimism (ENVOPT)	328
Table 16.29	Item parameters for responsibility for sustainable development (RESPDEV)	328
Table 16.30	Model fit environment-related constructs	329
Table 16.31	Estimated latent correlations for environment-related constructs	329
Table 16.32	Scale reliabilities for environment-related scales in OECD countries	330
Table 16.33	Scale reliabilities for environment-related scales in non-OECD countries	330
Table 16.34	Item parameters for school preparation for science career (CARPREP)	331
Table 16.35	Item parameters for student information on science careers (CARINFO)	331
Table 16.36	Model fit and estimated latent correlations for science career preparation indices	332
Table 16.37	Scale reliabilities for science career preparation indices	332
Table 16.38	Item parameters for science teaching: interaction (SCINTACT)	333
Table 16.39	Item parameters for science teaching: hands-on activities (SCHANDS)	333
Table 16.40	Item parameters for science teaching: student investigations (SCINVEST)	333
Table 16.41	Item parameters for science teaching: focus on models or applications (SCAPPLY)	334
Table 16.42	Model fit for CFA with science teaching and learning	334
Table 16.43	Estimated latent correlations for constructs related to science teaching and learning	335
Table 16.44	Scale reliabilities for scales to science teaching and learning in OECD countries	336
Table 16.45	Scale reliabilities for scales to science teaching and learning in partner countries/economies	336
Table 16.46	Item parameters for ICT Internet/entertainment use (INTUSE)	337
Table 16.47	Item parameters for ICT program/software use (PRGUSE)	337
Table 16.48	Item parameters for ICT self-confidence in Internet tasks (INTCONF)	337
Table 16.49	Item parameters for ICT self-confidence in high-level ICT tasks (HIGHCONF)	338
Table 16.50	Model fit for CFA with ICT familiarity items	338
Table 16.51	Estimated latent correlations for constructs related to ICT familiarity	339
Table 16.52	Scale reliabilities for ICT familiarity scales	339
Table 16.53	Item parameters for teacher shortage (TCSHORT)	340
Table 16.54	Item parameters for quality of educational resources (SCMATEDU)	340
Table 16.55	Item parameters for school activities to promote the learning of science (SCIPROM)	341
Table 16.56	Item parameters for school activities for learning environmental topics (ENVLEARN)	341
Table 16.57	Scale reliabilities for school-level scales in OECD countries	341
Table 16.58	Scale reliabilities for environment-related scales in partner countries/economies	342
Table 16.59	Item parameters for science activities at age 10 (PQSCIACT)	343
Table 16.60	Item parameters for parent's perception of school quality (PQSCHOOL)	343
Table 16.61	Item parameters for parent's views on importance of science (PQSCIMP)	343
Table 16.62	Item parameters for parent's reports on science career motivation (PQSCCAR)	344
Table 16.63	Item parameters for parent's view on general value of science (PQGENSCI)	344
Table 16.64	Item parameters for parent's view on personal value of science (PQPERSCI)	344
Table 16.65	Item parameters for parent's perception of environmental issues (PQENPERC)	345
Table 16.66	Item parameters for parent's environmental optimism (PQENVOPT)	345



Table 16.67	Scale reliabilities for parent questionnaire scales.....	345
Table 16.68	Factor loadings and internal consistency of ESCS 2006 in OECD countries.....	347
Table 16.69	Factor loadings and internal consistency of ESCS 2006 in partner countries/economies.....	348
<hr/>		
Table 17.1	Student-level latent correlations between mathematics, reading, science, embedded interest and embedded support.....	354
Table 17.2	Summary of the IRT scaling results across countries.....	355
Table 17.3	Gender DIF table for embedded attitude items.....	357
Table 17.4	Correlation amongst attitudinal scales, performance scales and HISEI.....	358
Table 17.5	Correlations for science scale.....	359
Table 17.6	Loadings of the achievement, interest and support variables on three varimax rotated components.....	360
Table 17.7	Correlation between embedded attitude scales and questionnaire attitude scales.....	361
Table 17.8	Rank order correlation five test domains, questionnaire attitude scales and HISEI.....	362
Table 17.9	Intra-class correlation (rho).....	362
<hr/>		
Table A1.1	2006 Main study reading item classification.....	380
Table A1.2	2006 Main study mathematics item classification.....	381
Table A1.3	2006 Main study science item classification (cognitive).....	383
Table A1.4	2006 Main study science embedded item classification (interest in learning science topics).....	387
Table A1.5	2006 Main study science embedded item classification (support for scientific enquiry).....	388
<hr/>		
Table A2.1	2006 Main study contrast coding used in conditioning for the student questionnaire variables.....	389
Table A2.2	2006 Main study contrast coding used in conditioning for the ICT questionnaire variables.....	396
Table A2.3	2006 Main study contrast coding used in conditioning for the parent questionnaire variables and other variables.....	397
<hr/>		
Table A3.1	Standard errors of the student performance mean estimate by country, by domain and cycle.....	399
Table A3.2	Sample sizes by country and cycle.....	400
Table A3.3	School variance estimate by country, by domain and cycle.....	401
Table A3.4	Intraclass correlation by country, by domain and cycle.....	402
Table A3.5	Within explicit strata intraclass correlation by country, by domain and cycle.....	403
Table A3.6	Percentages of school variance explained by explicit stratification variables, by domain and cycle.....	404
<hr/>		
Table A4.1	Student questionnaire.....	405
Table A4.2	ICT familiarity questionnaire.....	407
Table A4.3	School questionnaire.....	408
<hr/>		
Table A5.1	Mapping of ISCED to accumulated years of education.....	411
<hr/>		
Table A6.1	National household possession items.....	412
<hr/>		
Table A7.1	Exploratory and confirmatory factor analyses (EFA and CFA) for the embedded items.....	414



From:
PISA 2006 Technical Report

Access the complete publication at:
<https://doi.org/10.1787/9789264048096-en>

Please cite this chapter as:

OECD (2009), "Data Management Procedures", in *PISA 2006 Technical Report*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264048096-11-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.