



# Data Adjudication

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



## INTRODUCTION

This chapter describes the process used to adjudicate the implementation of PISA 2006 in each of the participating countries and adjudicated regions. It gives the outcomes of the data adjudication which are mainly based on the following aspects:

- The extent to which each country met PISA sampling standards;
- The outcomes of the adaptation, translation and verification processes;
- The outcomes of the national centre and PISA quality monitoring visits;
- The quality and completeness of the submitted data; and
- The outcomes of the international coding review.

In PISA 2006 all implementation procedures and documentations are developed in accordance with the Technical Standards.<sup>1</sup> The standards presented in that document were also used as the basis for data adjudication. The areas covered in those standards include the following:

- Target population and sampling:
  - Target population definitions, sample definitions, test period requirements;
  - School and student sampling response rates and coverage requirements;
  - Requirements for languages of assessment;
- Adaptation, translation and verification:
  - Adaptation of tests, questionnaires and manuals;
  - Translation of material and submission for translation and verification;
- Printing of materials;
- Common requirements for test administration procedures:
  - Selection and training of test administrators;
  - Security of material;
  - Conduct of testing sessions;
- Quality Monitoring:
  - Selection and training of PISA Quality Monitors (PQMs);
  - Site visits;
- Coding:
  - Single and multiple coding requirements;
  - International coding review;
- Data entry, processing and submission requirements.

### Implementing the standards – quality assurance

NPMs of participating countries and adjudicated regions are responsible for implementing the standards based on consortium advice as contained in the various operational manuals and guidelines. Throughout the cycle of activities for each PISA survey the consortium carried out quality assurance activities in two steps. The first step was to set up quality control using the operational manuals, as well as the agreement processes for national submissions on various aspects of the project. These processes give the consortium staff the opportunity to ensure that PISA implementation was planned in accordance with the PISA Technical Standards, and to provide advice on taking rectifying action when required and before critical errors occurred. The second step was quality monitoring, which involved the systematic collection of data that monitored the implementation of the assessment in relation to the standards. For data adjudication it was the information collected during both the quality control and quality monitoring activities that was used to determine the level of compliance with the standards.



## Information available for adjudication

The information collected by consortium staff during their quality control activities included communications and documentation exchanged with NPMs, and agreed national submissions which were stored on the PISA website. The quality monitoring instruments from which information was available included:

- PISA quality monitor reports;
- Test administrator session reports;
- Main study reviews;
- National centre quality monitor interview schedules.

Each of the quality monitoring instruments addressed different aspects of the standards and these were collected at different times during the data collection phase. There were two types of PQM reports, one containing data for each observed session in each school and another detailing the general observations across all schools visited by each quality monitor. The PQM reports contain data related to test administration as well as a record of interview with school coordinators. The test administrator session report was completed by the test administrator after each test session and also contained data related to test administration. The data from this report were data-entered by the national centre and submitted as part of the national dataset to the consortium. The *National Centre Quality Monitor Interview Schedule* contained information on all the standards, as did the *Main Study Review*.

The *National Centre Quality Monitor Interview Schedule* and the *Main Study Review* were self-declared by the NPM. The PQM data are collected independently of the NPM.

## Data adjudication process

The main aim of the adjudication process is to make a single determination on each national dataset in a manner that is transparent, based on evidence and defensible. The data adjudication process achieved this through the following steps:

- Step 1:** Quality control and quality monitoring data were collected throughout the survey cycle.
- Step 2:** Data collected from both quality control and quality monitoring activities were entered into a single quality assurance database.
- Step 3:** Experts compiled country-by-country reports that contained quality assurance data for key areas of project implementation.
- Step 4:** Experts considered the quality assurance data that were collected from both the quality control and quality monitoring activities, to make a judgement. In this phase the experts collaborated with the project director and other consortium staff to address any identified areas of concern. Where necessary, the relevant NPM was contacted through the project director. At the end of this phase experts constructed, for each adjudicated dataset, a summary detailing how the PISA technical standards had been met.
- Step 5:** The consortium and the Technical Advisory Group reviewed the reports and made a determination with regard to the quality of the data.

It was expected that the data adjudication would result in a range of possible recommendations. Some possible, foreseen recommendations included:

- That the data be declared fit for use;
- That some data be removed for a particular country, for example the removal of data for some items such as open-ended items, or the removal of data for some schools;



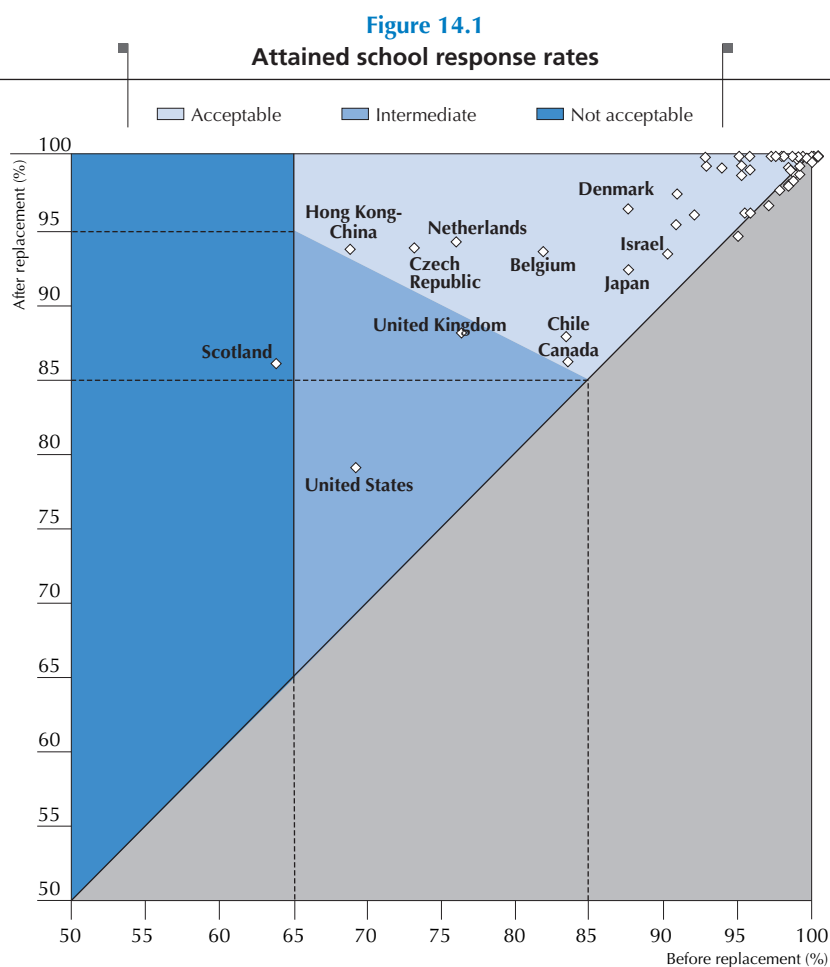
- That rectifying action be performed by the NPM, for example; providing additional evidence to demonstrate that there is no non-response bias, or rescoring open-ended items;
- That the data not be endorsed for use in certain types of analyses;
- That the data not be endorsed for inclusion in the PISA 2006 database.

Throughout PISA 2006 the consortium concentrated its quality control activities to ensure that the highest scientific standards were met. However during data adjudication a wider definition of quality was used especially when considering data that were at risk. In particular the underlying criterion used in adjudication was fitness for use. That is, data were endorsed for use if they were deemed to be fit for meeting the major intended purposes of PISA.

## GENERAL OUTCOMES

### Overview of response rate issues

The PISA school response rate requirements are discussed in Chapter 6. Figure 14.1 is a scatter plot of the attained PISA school response rates before and after replacements. Those countries that are plotted in the green shaded region were regarded as fully satisfying the PISA school response rate criterion.





Two countries – the United Kingdom (comprising two national centres, one to cover England, Wales and Northern Ireland; and Scotland which conducted the survey as a separate national centre) and the United States – failed to meet the school response rate requirements. In addition to failing the school response rate requirement, Scotland was the only participant to fail the student response rate requirement (see Chapter 11).

After reviewing the sampling outcomes, the consortium asked Scotland, the United Kingdom and the United States to provide additional data that would assist the consortium in making a balanced judgement about the threat of non-response to the accuracy of inferences that could be made from their PISA data.

### **Detailed country comments**

It is important to recognise that PISA data adjudication is a late but not necessarily final step in the quality assurance process. By the time each country was adjudicated at the TAG meeting that took place in Melbourne in March 2007, various quality assurance mechanisms (such as the sampling procedures documentation, translation verification, data cleaning and site visits) had already been applied at various stages of PISA 2006, and these had identified a range of issues. The purpose of these mechanisms was early identification of potential problems, and intervention to ensure that they had been rectified wherever possible so that data quality would be affected as little as possible. Details on the various quality assurance procedures and their outcomes are documented elsewhere (see Chapter 7).

Data adjudication focused on residual issues that remained after these quality assurance processes had been carried out. There were not many such issues and their projected impact on the validity of the PISA results was deemed to be negligible in most cases. These issues fall under two broad categories: 1) adaptations to the recommended international standard procedures in a country's data collection plan; and 2) a failure to meet international standards at the implementation stage.

#### ***Departures from standard procedures in the national data collection plan***

With such a broad and diverse range of participation, it is to be expected that the international best practice approaches to data collection articulated in the PISA Technical Standards document may not be achieved in all national and local contexts. This may be the case for a number of reasons. For example, it may be contrary to national protocols to have unannounced visits of quality monitors to schools to observe test administration. Or it may not be possible for teachers from very remote or very small schools to leave their schools to attend training in the mechanics of PISA test administration. Typically these were discussed with consortium experts in advance of the assessment and alternative approaches were considered jointly between the NPM and the consortium. In isolated departures from best practice in cases such as these, a judgement might easily be made by consortium experts that there was minimal risk in relation to the quality of the data collection plan. Such isolated departures are not reported in the country summaries below.

On the other hand, it may not have been straightforward to determine in advance of the assessment how more extensive, or multiple departures from PISA Standards may interact with each other, and with other aspects of a country's data collection plan. Cases such as these were considered as part of the data adjudication process, and are included in the country summaries below.

#### ***Departures from standards arising from implementation***

Departures from the standards at the implementation stage range from errors within the national centre (e.g. during the final stages of preparing materials, or in the administration of the coding operation following data collection), through to a failure to meet documented targets during data collection, for example a shortfall from the minimum school and student sample sizes.



A point in the preparation stage that led to significant errors in several countries was in the final stages of the preparation of the test booklets and questionnaire instruments at the national centre, following the final optical check of these materials by the international verification team (see Chapter 5). These errors included a failure to correct errors that had been identified by the international verifiers as part of the final optical check, or the introduction of completely new errors to the booklets and/or questionnaires following the final optical check. An obvious example of such an error (which was emphatically warned against, but nevertheless unfortunately occurred in a number of countries) is in the repagination of the booklets, so that the location of the item components (e.g. stimulus material and multiple-choice responses) would differ from the materials approved internationally. The nature and extent of such errors, the estimated impact on data quality, and actions taken with regard to the international database, are reported in the country summaries below.

A small number of countries failed to reach the required minimum sample sizes of 4500 students and 150 schools. Such cases were considered as part of the data adjudication process. Even a minor deviation in sample size might be considered a substantive enough issue to report, for example in countries where standard errors tend to be higher for a given sample size. On the other hand, minor deviations from these minimal sample sizes (*i.e.* shortfalls of fewer than 50 students or 5 schools, and in countries that nevertheless achieved comparable standard errors on the major survey estimates) are not reported below.

A component of the data adjudication process was to consider the cases of multiple, or more complex departures from the PISA standard procedures, as well as to consider the impact of errors or shortfalls across all aspects of each country's data collection plan and implementation, and make an evaluation with respect to the quality and international comparability of the PISA results. Notable departures from the standards are reported in the country summaries below. If a country is not listed below then it fully met the PISA standards. Further, in the case of minor deviations from the standards, unless otherwise noted, additional data was available to suggest the data was suitable for use.

### **Argentina**

Argentina had substantially fewer than the required 4 500 assessed students (4 297). More importantly Argentina had minor errors in their test booklet layout and there was a pagination error in the latter part of two of the 13 test booklets.

### **Azerbaijan**

The exclusion of occupied areas resulted in a coverage of 0.94 of the national enrolled population. There was also evidence of poor translation in some of the instruments at the field trial, which remained a concern at the main study. Many minor errors were observed in the administered test booklets, and print quality problems led to some re-printing of test materials. As a result of these issues the data from three items were deleted.

The Azerbaijan data was unusual in a number of regards. First the correlation between the Azerbaijan estimates of item difficulty and the international estimates is much lower than for any other participant. Second, the variation across booklets in student performance is far greater in Azerbaijan than is the case for any other participant. Third, the estimated variance in student mathematics performance in Azerbaijan was much less than for any other participant. Finally, as is frequently observed in low performing countries, there was an unusually high consistency among multiple coders.



### **Belgium**

A small percentage of TA's were trained by phone, rather than face-to-face and the TA reports show longer than expected session breaks.

#### ▪ **Flanders**

Inclusion of data from the Belgium region of Flanders for the adjudicated sub-national regions was recommended.

### **Brazil**

A high rate, 8.9%, of 'transferred' or not at school students was recorded.

### **Canada**

The overall level of exclusions was greater than 5% even after within-school language exclusions were removed. This high level of exclusions resulted in Canada's coverage of the national desired population and national enrolled population being 0.94 and 0.93 respectively.

### **Chile**

Some session timing irregularities in the testing sessions were observed, including extended breaks between sessions in four cases and 15 sessions of shorter duration than the standard.

The final weighted data contained only 46% female students. National statistics indicate that this figure should be about 49%. It was determined that this variation was explainable as sampling variance.

### **Colombia**

Pagination of the second half of three test booklets did not match the source version – these errors had been identified at FOC stage but were not rectified by the national centre.

### **Czech Republic**

Twelve schools used wrong instructions (affecting students using the UH booklet) and their data were discarded.

### **Denmark**

Overall exclusions were greater than 5% (6.07%), but just under 5% after within-school language exclusions were removed (4.96%). School level exclusions were greater than 2.5% (2.84%). Instructions for SEN exclusions were not correctly included in manuals.

### **Estonia**

School level exclusions were greater than 2.5%. The school exclusions from the initial sampling frame constituted 2.31% of the population, but exclusions identified in the field raised the school-level exclusion rate to 2.90%. The final exclusion rate from all sources was 3.97%, well within the PISA standard for overall exclusion. Thus it was determined that this slight violation of the PISA standards would have no appreciable impact on the quality or comparability of the data.

### **Finland**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.14%. There were errors in the printing of test booklets that resulted in some items being set to non-applicable for 30 students.



### **France**

The implementation of PISA in France deviated from the internationally recommended procedures in a number of ways. First, France did not implement the school questionnaire. It follows that France cannot be included in those reports and analyses that utilise school questionnaire data. Second, it was noted that the test administrators were not trained in person as required by the standards. As an alternative, the test administrators were trained through phone calls. Third, due to the exclusion of the *Territoires d'Outre-Mer* and to students in hospitals the French coverage of the national enrolled population was 0.93. Finally, due to local requirements, the PQMs were school inspectors and were not formally independent of the French national centre as was required by the standards.

### **Hungary**

School level exclusions were greater than 2.5%. The school exclusions from the initial sampling frame constituted 2.11% of the population, but exclusions identified in the field raised the school-level exclusion rate to 2.69%. The final exclusion rate from all sources was 3.69%, well within the PISA standard for overall exclusion. Thus it was determined that this slight violation of the PISA standards would have no appreciable impact on the quality or comparability of the data.

### **Iceland**

Test administrators were trained by phone. A small number of major item presentation errors were observed, including one item deletion due to a printing error.

### **Ireland**

Around one third of test sessions were reported as taking a break of more than 10 minutes between the two hours of the test session.

### **Israel**

Two reading items each in two booklets were set to not applicable due to item presentation and printing issues and seven items in the student questionnaire relating to responsibility for sustainable development were misprinted for all students; these items and the scale *RESPDEV* were set to not applicable.

### **Italy**

#### ▪ **Provincia Sicilia**

With a sample size of 1335, Sicilia had fewer than the required 1500 assessed students.

#### ▪ **Provincia Sardegna**

With a sample size of 1390, Sardegna had fewer than the required 1500 assessed students.

#### ▪ **Provincia Campania**

With a sample size of 1406, Campania had fewer than the required 1500 assessed students.

#### ▪ **Provincia Lombardia**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were at 2.12%.

### **Japan**

The implementation of PISA in Japan deviated from the internationally recommended procedures in a couple of ways. First, Japan had a high rate of absent students (6.4%); and second, all test administrators were the teachers of the students.

In the area of translation verification Japan implemented few of the key recommended corrections related to equivalence issues and the quality of the Japanese instruments was regarded as poor.



**Latvia**

School level exclusions were greater than 2.5%. The school exclusions from the initial sampling frame constituted 2.78% of the population. This was accepted by the consortium, as it was established that there would be no exclusion of special education students within school. The final exclusion rate from all sources was 3.21%, well within the PISA standard for overall exclusion.

**Luxembourg**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 0.51%.

**Macao-China**

Page layout for the English and Portuguese versions of the Macao test booklets did not match the international source versions; 9.4% of students responded to these booklets.

**Montenegro**

Montenegro had substantially fewer than the required 4500 assessed students (4367) and the coding guides were not submitted for final optical check.

**New Zealand**

Within-school exclusion rate was greater than 2.5% (3.84%) but after exclusions due to language were removed, it was 2.42%. The overall exclusion rate was 4.58%.

**Norway**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.04%. A small number of test administrators were trained by phone; 10% of TAs were teachers of the sampled students.

**Qatar**

Ten per cent of the students who were marked as present at the testing session wrote nothing in their test booklets. These students were treated as non-respondents. Pagination in eight of the thirteen Arabic booklets did not match the source versions and the parental occupation data were missing for around half the students.

**Slovak Republic**

A few pages in one test booklet were printed in the wrong order.

**Spain**

Within-school exclusions were greater than 2.5% (2.65%) but after exclusions due to language were removed, they were 1.73%. The overall exclusion rate was 3.52%.

All absent students were incorrectly coded as ineligible and this meant that student non-response adjustments could only be approximately calculated. No substantial bias is expected to have resulted from this. An additional consequence is that the population coverage rates cannot be correctly estimated. This error held for all of the adjudicated regions, so it is not listed for each case below.

**▪ Asturias**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.29%.



#### ▪ **Andalusia**

The sample size for Andalusia was 1463, slightly below the minimum requirement of 1500.

#### ▪ **Basque Country**

For the Spanish region of the Basque Country, the standard procedure relating to the language of assessment was not followed. In language settings involving the Basque language, students were tested in their home language rather than in their language of instruction (Basque). Note that as the Basque Country contains only a small percentage of the Spanish population this deviation does not influence the results for Spain overall.

In all other respects, the data for the Basque Country met the PISA standards. The consortium recommended that the Basque Country data be included in the full range of PISA reports and that the data be annotated where it is published to indicate that the PISA results in the Basque Country must be interpreted as the results obtained by the students enrolled in the Basque educational system, but not as the results obtained by the students attending instruction in Basque language.

#### ▪ **Cantabria**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were still 3.29%.

#### ▪ **Castile and Leon**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were still 3.03%.

#### ▪ **Catalonia**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.05%.

#### ▪ **La Rioja**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 1.75%. La Rioja had just 45 participating schools and a total sample size of 1335.

### **Sweden**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.13%. One mistranslation resulted in the responses to one item being set to not applicable.

### **Chinese Taipei**

The initial Chinese translations were not deemed satisfactory and it was suggested that Chinese Taipei adapt the version that the verifier had produced. Despite some improvements in the main study, the number of translation or adaptation problems in their final instruments remained high.

The sample of schools and students selected for the 2006 assessment covered only about 50% of the eligible population, with severe undercoverage of students in lower grades. Chinese Taipei undertook a follow-up assessment in 2007, in which a substantial sample of the previously non-covered population was assessed. The combination of these two samples provided fully satisfactory population coverage, met all PISA sampling standards, and was used for obtaining the final results.

### **Tunisia**

The print quality of the Tunisian test did not meet PISA standards.



### **United Kingdom**

The school response rate was 76.1% before replacement, and 88.2% after replacement. This placed the United Kingdom in the Intermediate zone for the school response rate standards. A major source of school non-response in the United Kingdom as a whole was from Scotland. The balance of the United Kingdom just missed meeting the PISA school response rate standard, and evidence concerning school non-response bias, supplied by the national centre, showed no evidence of bias. Given that the Scottish national centre provided evidence that there was no substantial school non-response bias for Scotland (see below), it was determined that there was no concern about significant school non-response bias for the United Kingdom as a whole.

#### ▪ **Scotland**

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were 2.26%.

The school response rate was 63.6% before replacement and therefore was in the not acceptable range. The national centre provided a detailed analysis of school non-response bias, which indicated no evidence of substantial bias resulting from school non-response. The response rate after replacement was 86.1%.

The student response rate was 78.6%, below the standard of 80%. The national centre provided a detailed analysis of student non-response bias. There was no evidence of substantial bias, based on the demographic characteristics of the respondents in comparison with the full sample. However, a substantial portion of the student non-response consisted of student refusals. This was markedly different from the case in previous PISA cycles for Scotland, meaning that some caution may be warranted in interpreting results related to trends over time.

### **United States**

The pagination of the test booklets did not match the source version. This error was introduced at the printing stage, after the consortium's final optical check. The pagination error was deemed to have invalidated the reading data, but its estimated effect on both mathematics and science was deemed to be negligible. The United States reading data were excluded from the database and international reports.

Within-school exclusions were greater than 2.5% but after exclusions due to language were removed, they were still 3.32%. The school response rate was 69.0% before replacement, and 79.1% after replacement, this placing the United States in the Intermediate zone for the school response rate standards. The National Centre provided a detailed analysis of school non-response bias, which indicated no evidence of substantial bias resulting from school non-response.





# 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](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/area06routitsky\\_berezner.pdf](https://mypisa.acer.edu.au/images/mypisadoc/area06routitsky_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<sup>®</sup> 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<sup>®</sup>: 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

<b>FOREWORD</b> .....	<b>3</b>
<b>CHAPTER 1 PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT: AN OVERVIEW</b> .....	<b>19</b>
<b>Participation</b> .....	21
<b>Features of PISA</b> .....	22
<b>Managing and implementing PISA</b> .....	23
<b>Organisation of this report</b> .....	23
<b>READER'S GUIDE</b> .....	<b>25</b>
<b>CHAPTER 2 TEST DESIGN AND TEST DEVELOPMENT</b> .....	<b>27</b>
<b>Test scope and format</b> .....	28
<b>Test design</b> .....	28
<b>Test development centres</b> .....	29
<b>Development timeline</b> .....	30
<b>The PISA 2006 scientific literacy framework</b> .....	30
<b>Test development – cognitive items</b> .....	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
<b>Test development – attitudinal items</b> .....	35
<b>Field trial</b> .....	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
<b>Main study</b> .....	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



<b>CHAPTER 3 THE DEVELOPMENT OF THE PISA CONTEXT QUESTIONNAIRES</b> .....	<b>49</b>
<b>Overview</b> .....	50
<b>The conceptual structure</b> .....	51
▪ A conceptual framework for PISA 2006 .....	51
<b>Research areas in PISA 2006</b> .....	55
<b>The development of the context questionnaires</b> .....	57
<b>The coverage of the questionnaire material</b> .....	58
▪ Student questionnaire .....	58
▪ School questionnaire .....	59
▪ International options .....	59
▪ National questionnaire material .....	60
<b>The implementation of the context questionnaires</b> .....	60
 <b>CHAPTER 4 SAMPLE DESIGN</b> .....	 <b>63</b>
<b>Target population and overview of the sampling design</b> .....	64
<b>Population coverage, and school and student participation rate standards</b> .....	65
▪ Coverage of the PISA international target population .....	65
▪ Accuracy and precision .....	66
▪ School response rates .....	66
▪ Student response rates .....	68
<b>Main study school sample</b> .....	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
 <b>CHAPTER 5 TRANSLATION AND CULTURAL APPROPRIATENESS OF THE TEST AND SURVEY MATERIAL</b> .....	 <b>85</b>
<b>Introduction</b> .....	86
<b>Development of source versions</b> .....	86
<b>Double translation from two source languages</b> .....	87
<b>PISA translation and adaptation guidelines</b> .....	88
<b>Translation training session</b> .....	89
<b>Testing languages and translation/adaptation procedures</b> .....	89
<b>International verification of the national versions</b> .....	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



<b>Translation and verification outcomes – national version quality</b> .....	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
<b>CHAPTER 6 FIELD OPERATIONS</b> .....	<b>105</b>
<b>Overview of roles and responsibilities</b> .....	106
▪ National project managers.....	106
▪ School coordinators.....	107
▪ Test administrators.....	107
▪ School associates.....	108
<b>The selection of the school sample</b> .....	108
<b>Preparation of test booklets, questionnaires and manuals</b> .....	108
<b>The selection of the student sample</b> .....	109
<b>Packaging and shipping materials</b> .....	110
<b>Receipt of materials at the national centre after testing</b> .....	110
<b>Coding of the tests and questionnaires</b> .....	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
<b>Data entry, data checking and file submission</b> .....	120
▪ Data entry.....	120
▪ Data checking.....	120
▪ Data submission.....	121
▪ After data were submitted.....	121
<b>The main study review</b> .....	121
<b>CHAPTER 7 QUALITY ASSURANCE</b> .....	<b>123</b>
<b>PISA quality control</b> .....	124
▪ Comprehensive operational manuals.....	124
▪ National level implementation planning document.....	124
<b>PISA quality monitoring</b> .....	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
<b>CHAPTER 8 SURVEY WEIGHTING AND THE CALCULATION OF SAMPLING VARIANCE</b> .....	<b>129</b>
<b>Survey weighting</b> .....	130
<b>The school base weight</b> .....	131
▪ The school weight trimming factor.....	132

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



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



<b>CHAPTER 14 DATA ADJUDICATION</b> .....	<b>271</b>
<b>Introduction</b> .....	272
▪ Implementing the standards – quality assurance .....	272
▪ Information available for adjudication .....	273
▪ Data adjudication process .....	273
<b>General outcomes</b> .....	274
▪ Overview of response rate issues .....	274
▪ Detailed country comments .....	275
<b>CHAPTER 15 PROFICIENCY SCALE CONSTRUCTION</b> .....	<b>283</b>
<b>Introduction</b> .....	284
<b>Development of the described scales</b> .....	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
<b>Defining proficiency levels</b> .....	287
<b>Reporting the results for PISA science</b> .....	290
▪ Building an item map .....	290
▪ Levels of scientific literacy .....	292
▪ Interpreting the scientific literacy levels .....	299
<b>CHAPTER 16 SCALING PROCEDURES AND CONSTRUCT VALIDATION OF CONTEXT QUESTIONNAIRE DATA</b> .....	<b>303</b>
<b>Overview</b> .....	304
<b>Simple questionnaire indices</b> .....	304
▪ Student questionnaire indices .....	304
▪ School questionnaire indices .....	307
▪ Parent questionnaire indices .....	309
<b>Scaling methodology and construct validation</b> .....	310
▪ Scaling procedures .....	310
▪ Construct validation .....	312
▪ Describing questionnaire scale indices .....	314
<b>Questionnaire scale indices</b> .....	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
<b>CHAPTER 17 VALIDATION OF THE EMBEDDED ATTITUDINAL SCALES</b> .....	<b>351</b>
<b>Introduction</b> .....	352
<b>International scalability</b> .....	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
<b>Relationship and comparisons with other variables.....</b>	<b>357</b>
▪ 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
<b>Conclusion.....</b>	<b>364</b>
<b>CHAPTER 18 INTERNATIONAL DATABASE.....</b>	<b>367</b>
<b>Files in the database.....</b>	<b>368</b>
▪ Student files.....	368
▪ School file.....	370
▪ Parent file.....	370
<b>Records in the database.....</b>	<b>371</b>
▪ Records included in the database.....	371
▪ Records excluded from the database.....	371
<b>Representing missing data.....</b>	<b>371</b>
<b>How are students and schools identified?.....</b>	<b>372</b>
<b>Further information.....</b>	<b>373</b>
<b>REFERENCES.....</b>	<b>375</b>
<b>APPENDICES.....</b>	<b>379</b>
<b>Appendix 1</b> PISA 2006 main study item pool characteristics.....	380
<b>Appendix 2</b> Contrast coding used in conditioning.....	389
<b>Appendix 3</b> Design effect tables.....	399
<b>Appendix 4</b> Changes to core questionnaire items from 2003 to 2006.....	405
<b>Appendix 5</b> Mapping of ISCED to years.....	411
<b>Appendix 6</b> National household possession items.....	412
<b>Appendix 7</b> Exploratory and confirmatory factor analyses for the embedded items.....	414
<b>Appendix 8</b> 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



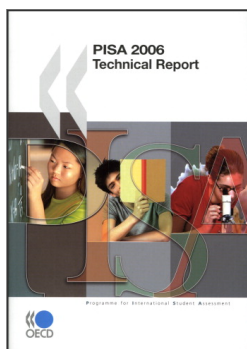
Table 11.11	Effective sample size 3 by country, by domain and cycle .....	198
Table 11.12	Design effect 4 by country, by domain and cycle.....	199
Table 11.13	Effective sample size 4 by country, by domain and cycle .....	200
Table 11.14	Design effect 5 by country, by domain and cycle.....	201
Table 11.15	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 Adjudication", in *PISA 2006 Technical Report*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264048096-15-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](mailto: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](mailto:info@copyright.com) or the Centre français d'exploitation du droit de copie (CFC) at [contact@cfcopies.com](mailto:contact@cfcopies.com).