PISA 2012
Technical Report

Programme for International Student Assessment

OECD

PISA 2012
Technical Report

OECD
BETTER POLICIES FOR BETTER LIVES

Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Photo credits:
© Flying Colours Ltd/Getty Images
© Jacobs Stock Photography/Kzenon
© khoa vu/Flickr/Getty Images
© Mel Curtis/Corbis
© Shutterstock/Kzenon
© Simon Jarratt/Corbis

Foreword

The OECD Programme for International Student Assessment (PISA) surveys, which take place every three years, have been designed to collect information about 15-year-old students in participating countries. PISA examines how well students are prepared to meet the challenges of the future, rather than how well they master particular curricula. The data collected during each PISA cycle are an extremely valuable source of information for researchers, policy makers, educators, parents and students. It is now recognised that the future economic and social well-being of countries is closely linked to the knowledge and skills of their populations. The internationally comparable information provided by PISA allows countries to assess how well their 15-year-old students are prepared for life in a larger context and to compare their relative strengths and weaknesses.

PISA is methodologically highly complex, requiring intensive collaboration among many stakeholders. The successful implementation of PISA depends on the use, and sometimes further development, of state-of-the-art methodologies and technologies. The *PISA 2012 Technical Report* describes those methodologies, along with other features that have enabled PISA to provide high quality data to support policy formation and review. The descriptions are provided at a level that will enable review and, potentially, replication of the implemented procedures and technical solutions to problems.

This report contains a description of the theoretical underpinning of the complex techniques used to create the *PISA 2012 database*, which includes information on 510 000 students in 65 countries or economies. The database includes not only information on student performance in the main areas of assessment – mathematics, reading, science, problem solving and financial literacy – but also their responses to the Student Questionnaire that they completed as part of the assessment. Data from the principals of participating schools are also included. The PISA 2012 database was used to generate information and to be the basis for analysis for the PISA 2012 initial report, *PISA 2012 Results* (OECD, 2013 and 2014).

The information in this report complements the *PISA Data Analysis Manuals* (OECD, 2009), which give detailed accounts of how to carry out the analyses of the information in the database.

The PISA surveys are guided by the governments of the participating countries on the basis of shared policy-driven interests. The PISA Governing Board, which decides on the assessment and reporting of results, is composed of representatives from each participating country and economy.

The OECD recognises the creative work of Raymond Adams, of the Australian Council for Educational Research (ACER), who is project director of the PISA Consortium, and Ross Turner who acted as editor for this report. The team supporting them comprised Susan Bates, Alla Berezner, Jonas Bertling, Renee Chow, John Cresswell, Alexander Daraganov, Steve Dept, Andrea Ferrari, Béatrice Halleux, Eckhard Klieme, Nora Kovarcikova, Sheila Krawchuk, Petra Lietz, Greg Macaskill, Juliette Mendelovits, Alla Routitsky, Keith Rust, Stephanie Templeton and Maurice Walker. Christian Monseur and Maciej Jakubowski provided technical advice to review this report. A full list of the contributors to the PISA project is included in Annex G of this report. The editorial work at the OECD Secretariat was carried out by Josefa Palacios, Giannina Rech, Sophie Vayssettes and Élisabeth Villoutreix.

Lorna Bertrand

Lorna Bertrand
Chair of the PISA Governing Board

Andreas Schleicher

Andreas Schleicher
Director for Education and Skills, OECD

Table of Contents

BOXES

FIGURES

TABLES

Reader's Guide

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

ACER:	Australian Council for Educational Research
aSPe:	University of Liege, Belgium
BAS:	Booklet Adaptation Spreadsheet
BRR:	Balanced Repeated Replication
CBA:	Computer-based assessment
CITO:	National Institute for Educational Measurement, the Netherlands
DIF:	Differential Item Functioning
DIPF:	The German Institute for International Educational Research
DRA:	Digital Reading Assessment
DTCS:	DRA Target Cluster Size
ENR:	Enrolment of 15-year-olds
ESCS:	PISA Index of educational, social and cultural status
ETCS:	CBA Tagert Cluster Size
ETS:	Educational Testing Service
FOC:	Final Optical Check
FT:	Field Trial
I:	Sampling interval
ICF:	Item Characteristic Function
ICR:	International Coding Review
ICT:	Information Communication Technology
IEA:	International Association for the Evaluation of Educational Achievement
ILS:	University of Oslo, Norway
INES:	OECD Indicators of Education Systems
INT:	International
IPN:	Leibniz Institute for Science and Mathematics Education, Germany
IRT:	Item Response Theory
ISCED:	International Standard Classification of Education
ISCO:	International Standard Classification of Occupations
ISEI:	International Socio-Economic Index
MAS:	Manuals Adaptation Spreadsheets
MEG:	Mathematics Expert Group
MENR:	Enrolment for moderately small school
MNSQ:	Mean square
MOS:	Measure of size
MS:	Main Survey
NCQM:	National Centre Quality Monitor
NEP:	National enrolled population
NIER:	National Institute for Educational Research, Japan
NPM:	National Project Manager
OLT:	Open Language Tool
OTL:	Opportunity To Learn
PBA:	Paper-based assessment
PCA:	Principal Component Analysis
PGB:	PISA Governing Board
PPS:	Probability Proportional to Size
PQM:	PISA Quality Monitor
PV:	Plausible Values
QAS:	Questionnaire Adaptations Spreadsheet
RN:	Random Number
RP:	Response Probability
SC:	School Co-ordinator
S.D.:	Standard Deviation
S.E.:	Standard Error
SEN:	Special Education Needs
SJT:	Situational Judgment Tests
TA:	Test Administrator
TAG:	Technical Advisory Group
TAS:	Test Adaptation Spreadsheet
TCS:	Target Cluster Size
TIMSS:	Third International Mathematics and Science Study
TMS:	Translation Management System
UH:	Une Heure booklet
WLE:	Weighted Likelihood Estimates

List of country codes – the following country codes are used in some tables in this report:

Country/Economy	ISO code	Country/Economy	ISO code
Albania	ALB	Kazakhstan	KAZ
United Arab Emirates	ARE	Korea	KOR
Argentina	ARG	Liechtenstein	LIE
Australia	AUS	Lithuania	LTU
Austria	AUT	Luxembourg	LUX
Belgium	BEL	Latvia	LVA
Bulgaria	BGR	Macao-China	MAC
Brazil	BRA	Mexico	MEX
Canada	CAN	Montenegro	MNE
Switzerland	CHE	Malaysia	MYS
Chile	CHL	Netherlands	NLD
Colombia	COL	Norway	NOR
Costa Rica	CRI	New Zealand	NZL
Cyprus[1, 2]	CYP	Peru	PER
Czech Republic	CZE	Poland	POL
Germany	DEU	Portugal	PRT
Denmark	DNK	Qatar	QAT
Spain	ESP	Shanghai-China	QCN
Estonia	EST	Romania	ROU
Finland	FIN	Russian Federation	RUS
France	FRA	Singapore	SGP
United Kingdom	GBR	Serbia	SRB
Greece	GRC	Slovak Republic	SVK
Hong Kong-China	HKG	Slovenia	SVN
Croatia	HRV	Sweden	SWE
Hungary	HUN	Chinese Taipei	TAP
Indonesia	IDN	Thailand	THA
Ireland	IRL	Tunisia	TUN
Iceland	ISL	Turkey	TUR
Israel	ISR	Uruguay	URY
Italy	ITA	United States	USA
Jordan	JOR	Viet Nam	VNM
Japan	JPN		

1. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

2. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

The reader should note that a series of technical documents are available from the PISA website: *www.oecd.org/pisa*.

1

Programme for International Student Assessment: an Overview

The OECD Programme for International Student Assessment (PISA) is a collaborative effort among OECD member countries to measure how well 15-year-old students approaching the end of compulsory schooling are prepared to meet the challenges of today's knowledge societies. The assessment is forward-looking: rather than focusing on the extent to which these students have mastered a specific school curriculum, it looks at their ability to use their knowledge and skills to meet real-life challenges. This orientation reflects a change in curricular goals and objectives, which are increasingly concerned with what students can do with what they learn at school.

PISA surveys take place every three years. The first survey took place in 2000 (followed by a further 11 countries in 2002), the second in 2003, the third in 2006, the fourth in 2009 (followed by a further 10 countries and economies in 2010), and the fifth in 2012; the results of these surveys have been published in a series of reports (OECD, 2001, 2004, 2007, 2010, 2011, 2013, 2014) and a wide range of thematic and technical reports. The next survey will occur in 2015. For each assessment, one of reading, mathematics and science is chosen as the major domain and given greater emphasis. The remaining two areas, the minor domains, are assessed less thoroughly. In 2000 and 2009 the major domain was reading; in 2003 and 2012 it was mathematics and in 2006 it was science.

PISA is an age-based survey, assessing 15-year-old students in school in grade 7 or higher. These students are approaching the end of compulsory schooling in most participating countries, and school enrolment at this level is close to universal in almost all OECD countries.

The PISA assessments take a literacy perspective, which focuses on the extent to which students can apply the knowledge and skills they have learned and practised at school when confronted with situations and challenges for which that knowledge may be relevant. That is, PISA assesses the extent to which students can use their reading skills to understand and interpret the various kinds of written material that they are likely to meet as they negotiate their daily lives; the extent to which students can use their mathematical knowledge and skills to solve various kinds of numerical and spatial challenges and problems; and the extent to which students can use their scientific knowledge and skills to understand, interpret and resolve various kinds of scientific situations and challenges. The PISA 2012 domain definitions are fully articulated in *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013a).

PISA also allows for the assessment of additional cross-curricular competencies from time to time as participating countries see fit. For example, in PISA 2003, an assessment of general problem-solving competencies was included. A major addition for PISA 2009 was the inclusion of a computer-delivered assessment of digital reading which is also known as the digital reading assessment (DRA). For 2012 a computer-delivered assessment of mathematics and problem solving was added, along with an assessment of financial literacy.

PISA also uses Student Questionnaires to collect information from students on various aspects of their home, family and school background, and School Questionnaires to collect information from schools about various aspects of organisation and educational provision in schools. In PISA 2012, 11 countries also administered a Parent Questionnaire to the parents of the students participating in PISA.

Using the data from Student, Parent and School Questionnaires, analyses linking contextual information with student achievement could address:

- differences between countries in the relationships between student-level factors (such as gender and socio-economic background) and achievement;
- differences in the relationships between school-level factors and achievement across countries;
- differences in the proportion of variation in achievement between (rather than within) schools, and differences in this value across countries;
- differences between countries in the extent to which schools moderate or increase the effects of individual-level student factors and student achievement;
- differences in education systems and national context that are related to differences in student achievement across countries; and
- through links to PISA 2000, PISA 2003, PISA 2006 and PISA 2009, changes in any or all of these relationships over time.

Through the collection of such information at the student and school level on a cross-nationally comparable basis, PISA adds significantly to the knowledge base that was previously available from national official statistics, such as aggregate national statistics on the educational programmes completed and the qualifications obtained by individuals.

The framework for the PISA 2012 questionnaires is included in *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013a).

PARTICIPATION

The first PISA survey was conducted in 2000 in 32 countries (including 28 OECD member countries) using written tasks answered in schools under independently supervised test conditions. Another 11 countries completed the same assessment in 2002. PISA 2000 surveyed reading, mathematics and science, with a primary focus on reading.

The second PISA survey, conducted in 2003 in 41 countries, assessed reading, mathematics and science, and problem solving with a primary focus on mathematics. The third survey covered reading, mathematics and science, with a primary focus on science, and was conducted in 2006 in 57 countries.

PISA 2009, the fourth PISA survey covered reading, mathematics and science, with a primary focus on reading, and was conducted in 65 countries and economies. Another 10 additional participants completed the PISA 2009 assessment in 2010.

PISA 2012, the fifth PISA survey covered reading, mathematics, science, problem solving and financial literacy with a primary focus on mathematics, and was conducted in 34 OECD countries and 31 partner countries/economies. The participants in PISA 2012 are listed in Figure 1.1. The figure also indicates the 44 countries/economies that participated in the computer-delivered assessment of problem solving, the 32 countries/economies who participated in the computer-based assessment of mathematics and reading, and the 18 countries/economies who participated in the assessment of financial literacy.

This report is concerned with the technical aspects of PISA 2012.

■ Figure 1.1 [Part 1/2] ■

PISA 2012 participants

OECD countries	Computer-based assessment of mathematics and reading	Problem solving	Financial literacy
Australia	Yes	Yes	Yes
Austria	Yes	Yes	No
Belgium	Yes	Yes	Yes[1]
Canada	Yes	Yes	No
Chile	Yes	Yes	No
Czech Republic	No	Yes	Yes
Denmark	Yes	Yes	No
Estonia	Yes	Yes	Yes
Finland	No	Yes	No
France	Yes	Yes	Yes
Germany	Yes	Yes	No
Greece	No	No	No
Hungary	Yes	Yes	No
Iceland	No	No	No
Ireland	Yes	Yes	No
Israel	Yes	Yes	Yes
Italy	Yes	Yes	Yes
Japan	Yes	Yes	No
Korea	Yes	Yes	No
Luxembourg	No	No	No
Mexico	No	No	No
Netherlands	No	Yes	No
New Zealand	No	No	Yes
Norway	Yes	Yes	No
Poland	Yes	Yes	Yes
Portugal	Yes	Yes	No
Slovak Republic	Yes	Yes	Yes
Slovenia	Yes	Yes	Yes
Spain	Yes	Yes	Yes
Sweden	Yes	Yes	No
Switzerland	No	No	No
Turkey	No	Yes	No
United Kingdom	No	Yes[2]	No
United States	Yes	Yes	Yes

1. Only the Flemish Community of Belgium participated in the financial literacy assessment.
2. Only England participated in the problem-solving assessment.

■ Figure 1.1 [Part 2/2] ■

PISA 2012 participants

Partner countries/economies	Computer-based assessment of mathematics and reading	Problem solving	Financial literacy
Albania	No	No	No
Argentina	No	No	No
Brazil	Yes	Yes	No
Bulgaria	No	Yes	No
Colombia	Yes	Yes	Yes
Costa Rica	No	No	No
Croatia	No	Yes	Yes
Cyprus*	No	Yes	No
Hong Kong-China	Yes	Yes	No
Indonesia	No	No	No
Jordan	No	No	No
Kazakhstan	No	No	No
Latvia	No	No	Yes
Liechtenstein	No	No	No
Lithuania	No	No	No
Macao-China	Yes	Yes	No
Malaysia	No	Yes	No
Montenegro	No	Yes	No
Peru	No	No	No
Qatar	No	No	No
Romania	No	No	No
Russian Federation	Yes	Yes	Yes
Serbia	No	Yes	No
Shanghai-China	Yes	Yes	Yes
Singapore	Yes	Yes	No
Chinese Taipei	Yes	Yes	No
Thailand	No	No	No
Tunisia	No	No	No
United Arab Emirates	Yes	Yes	No
Uruguay	No	Yes	No
Viet Nam	No	No	No
Total	**32**	**44**	**18**

1. Only the Flemish Community of Belgium participated in the financial literacy assessment.

2. Only England participated in the problem-solving assessment.

* Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

FEATURES OF PISA

The technical characteristics of the PISA survey involve a number of different aspects:

- the design of the test and the features incorporated into the test developed for PISA are critical;
- the sampling design, including both the school sampling and the student sampling requirements and procedures;
- because of the multilingual nature of the test, rules and procedures are required to guarantee the equivalence of the different language versions used within and between participating countries, and to take into account the diverse cultural contexts of those countries;
- various operational procedures, including test administration arrangements, data capture and processing and quality assurance mechanisms designed to ensure the generation of comparable data from all countries; and
- scaling and analysis of the data and their subsequent reporting: PISA employs scaling models based on Item Response Theory (IRT) methodologies. The described proficiency scales, which are the basic tool in reporting PISA outcomes, are derived using IRT analysis.

This report describes the above-mentioned methodologies as they have been implemented in PISA 2012. It also describes the quality assurance procedures that have enabled PISA to provide high quality data to support policy formation and review. Box 1.1 provides an overview of the central design elements of PISA 2012.

The ambitious goals of PISA come at a cost: PISA is both resources intensive and methodologically complex, requiring intensive collaboration among many stakeholders. The successful implementation of PISA depends on the use, and sometimes further development, of state-of-the-art methodologies.

Quality within each of these areas is defined, monitored and assured through the use of a set of technical standards. These standards have been endorsed by the PISA Governing Board, and they form the backbone of implementation in each participating country and of quality assurance across the project (see Annex F for the PISA 2012 Technical Standards).

Box 1.1. **Key features of PISA 2012**

The content

The PISA 2012 survey focused on mathematics, with reading, science and problem-solving as minor areas of assessment. For the first time, PISA 2012 also included an assessment of the financial literacy of young people, which was optional for countries and economies.

PISA assesses not only whether students can reproduce knowledge, but also whether they can extrapolate from what they have learned and apply their knowledge in new situations. It emphasises the mastery of processes, the understanding of concepts, and the ability to function in various types of situations.

The students

Around 510 000 students completed the assessment in 2012, representing about 28 million 15-year-olds in the schools of the 65 participating countries and economies.

The assessment

Paper-based tests were used, with assessments lasting a total of two hours for each student. In a range of countries and economies, an additional 40 minutes were devoted to the computer-based assessment of mathematics, reading and problem solving.

Test items were a mixture of multiple-choice items and questions requiring students to construct their own responses. The items were organised in groups based on a text or graphic setting out a real-life situation. A total of about 390 minutes of test items was included, with different students taking different combinations of test items.

Students answered a background questionnaire, which took 30 minutes to complete, that sought information about themselves, their homes and their school and learning experiences. School principals were given a questionnaire, to complete, that covered the school system and the learning environment. In some countries and economies, optional questionnaires were distributed to parents, who were asked to provide information on their perceptions of and involvement in their child's school, their support for learning in the home, and their child's career expectations, particularly in mathematics-based occupations. Countries could choose two other optional questionnaires for students: one asked students about their familiarity with and use of information and communication technologies, and the second sought information about their education to date, including any interruptions in their schooling and whether and how they are preparing for a future career.

MANAGING AND IMPLEMENTING PISA

The design and implementation of PISA for the 2000, 2003, 2006, 2009 and 2012 data collections was the responsibility of an international consortium led by the Australian Council for Educational Research (ACER) with Ray Adams as International Project Director. Achieve (United States) was contracted by the OECD to develop the mathematics framework with ACER.

For PISA 2012 the Consortium partners were:

- cApStAn Linguistic Quality Control (Belgium)
- Deutsches Institut für Internationale Pädagogische Forschung (DIPF, Germany)

- Educational Testing Service (ETS, United States)
- Institutt for Lærerutdanning og Skoleutvikling (ILS, Norway)
- Leibniz - Institute for Science and MathematicsEducation (IPN, Germany)
- National Institute for Educational Policy Research (NIER, Japan)
- The Tao Initiative: CRP - Henri Tudor and Université de Luxembourg - EMACS (Luxembourg)
- Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe, Belgium)
- Westat (United States)

Annex G lists the Consortia staff and consultants who have made significant contributions to the development and implementation of the project.

PISA is implemented within a framework established by the PISA Governing Board (PGB) which includes representation from all participating countries at senior policy levels. The PGB established policy priorities and standards for developing indicators, for establishing assessment instruments, and for reporting results. Experts from participating countries served on working groups linking the programme policy objectives with the best internationally available technical expertise in the three assessment areas and in the areas which were included in the context questionnaires.

These expert groups were referred to as Subject Matter Expert Groups (EGs) (see Annex G for the list of members) and the Questionnaire Expert Group (QEG). By participating in these expert groups and regularly reviewing outcomes of the groups' meetings, countries ensured that the instruments were internationally valid, that they took the cultural and educational contexts of the different OECD member countries into account, that the assessment materials had strong measurement potential, and that the instruments emphasised authenticity and educational validity.

Each of the participating countries appointed a National Project Manager (NPM), to implement PISA nationally. The NPMs ensured that internationally agreed common technical and administrative procedures were employed. These managers played a vital role in developing and validating the international assessment instruments and ensured that PISA implementation was of high quality. The NPMs also contributed to the verification and evaluation of the survey results, analyses and reports.

The OECD Secretariat was responsible for the overall management of the programme. It monitored its implementation on a day-to-day basis, served as the secretariat for the PGB, fostered consensus building between the countries involved, and served as the interlocutor between the PGB and the international Consortia.

STRUCTURE OF THIS REPORT

This Technical Report is designed to describe the technical aspects of the project at a sufficient level of detail to enable review and, potentially, replication of the implemented procedures and technical solutions to problems. It therefore does not report the results of PISA 2012 which have been published in *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I – Revised Edition)* (OECD, 2014a), *Excellence through Equity: Giving Every Student the Chance to Succeed (Volume II)* (OECD, 2013b), *Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)* (OECD, 2013c), *What Makes Schools Successful? Resources, Policies and Practices (Volume IV)* (OECD, 2013d), *Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V)* (OECD, 2014b), *Students and Money: Financial Literacy Skills for the 21st century (Volume VI*) (OECD, 2014c). A bibliography of other PISA related reports is included in Annex H.

There are five sections in this report:

- *Section One — Instrument design:* describes the design and development of both the questionnaires and achievement tests (Chapters 2 and 3).
- *Section Two — Operations:* gives details of the operational procedures for the sampling and population definitions, test administration procedures, quality monitoring and assurance procedures for Test Administration and National Centre operations, and instrument translation (Chapters 4, 5, 6 and 7).
- *Section Three — Data processing:* covers the methods used in data cleaning and preparation, including the methods for weighting and variance estimation, scaling methods, methods for examining inter-rater variation and the data cleaning steps (Chapters 8, 9 and 10).

- *Section Four — Quality indicators and outcomes:* covers the results of the scaling and weighting, report response rates and related sampling outcomes and gives the outcomes of the inter-rater reliability studies. The last chapter in this section summarises the outcomes of the PISA 2012 data adjudication; that is, the overall analysis of data quality for each country (Chapters 11, 12, 13 and 14).
- *Section Five — Scale construction and data products:* describes the construction of the PISA 2012 described levels of proficiency and the construction and validation of questionnaire-related indices. The final chapter briefly describes the contents of the PISA 2012 database (Chapters 15, 16, 17, 18 and 19).

There are also detailed annexes of results pertaining to the chapters of the report that are provided.

References

OECD (2014a), *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

OECD (2014b), *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208070-en

OECD (2014c), *PISA 2012 Results: Students and Money (Volume VI): Financial Literacy Skills for the 21st Century,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208094-en

OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264190511-en

OECD (2013b), *PISA 2012 Results: Excellence through Equity (Volume II): Giving Every Student the Chance to Succeed,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201132-en

OECD (2013c), *PISA 2012 Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-beliefs,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201170-en

OECD (2013d), *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201156-en

OECD (2011), *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264112995-en

OECD (2010a), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264091450-en

OECD (2010b), *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes (Volume II),* PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264091504-en

OECD (2010c), *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices (Volume III),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264083943-en

OECD (2010d), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264091559-en

OECD (2010e), *PISA 2009 Results: Learning Trends: Changes in Student Performance Since 2000 (Volume V),* PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264091580-en

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World,* PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264040014-en

OECD (2004), *Learning for Tomorrow's World: First Results from PISA 2003*, PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264006416-en

OECD (2001), *Knowledge and Skills for Life: First Results from PISA 2000*, PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264195905-en

2

Test Design and Test Development

This chapter describes the test design for PISA 2012 and the processes by which the PISA Consortium, led by the Australian Council for Educational Research (ACER), developed the PISA 2012 paper-based tests for mathematics, reading and science, as well as for the international option, financial literacy. It also describes the design and development of the computer-based assessments of problem solving, mathematics and reading. In the following discussion, the term "mathematics" generally refers to the core paper-based mathematics assessment. The computer-based assessment of mathematics is referred to as "CBAM". The same applies in the case of reading: the computer-based assessment is referred to as the "digital reading assessment" (DRA). The PISA results reported publicly in December 2013 were from what is referred to as the PISA "Main Survey". This term is used to distinguish earlier developmental activities including those contributing to conduct of the "Field Trial" that occurred in 2011.

TEST SCOPE, DESIGN AND DEVELOPMENT

Test development for the PISA 2012 survey commenced in late 2009. Development proceeded through various processes and stages, slightly different for each of the cognitive domains in which test material was required, and culminating in the presentation to the PISA Governing Board (PGB) in October 2011 of a selection of items proposed for use in the 2012 Main Survey. This chapter presents the test design that governed the scope and structure of the PISA 2012 assessment, the development arrangements and approaches taken by ACER to produce the material required, and the processes of test development in each domain. Those domain-specific processes commenced with the specifications laid out in each assessment framework, and proceeded through the various stages of soliciting material for consideration, developing and refining that material to a finished form, seeking national feedback on the item developed, piloting and trialing draft material, and preparing materials fit and ready for use in the Main Survey.

The test design adopted for PISA 2012 specified the volume and arrangement of material needed in each domain that was to be tested (mathematics, reading, science, problem solving and financial literacy), and in each test mode that was to be employed (paper-based and computer-based). Those specifications required the development of sets of items (referred to as "item clusters") in each test domain, each of which would need to occupy a defined amount of test time. The specifications also determined how the item clusters would be arranged in test booklets (for the paper-based components) and in test forms (for the computer-based components).

Paper-based assessment design: mathematics, reading, science, financial literacy

The standard Main Survey items for mathematics, reading and science were to be compiled in thirteen item clusters (seven mathematics clusters, three reading clusters and three science clusters) with each cluster representing 30 minutes of test time. The items were presented to students in thirteen standard test booklets, with each booklet being composed of four clusters, hence two hours of test time. Clusters labelled PM1, PM2, PM3, PM4, PM5, PM6A and PM7A denote the seven paper-based standard mathematics clusters, PR1 to PR3 denote the paper-based reading clusters, and PS1 to PS3 denote the paper-based science clusters.

PM1, PM2 and PM3 were the same three mathematics clusters as those administered in 2009, and the remaining clusters would comprise new material. Two of the three reading clusters were intact clusters used in 2009. The remaining reading cluster was based on a cluster used in 2009 but with one unit substituted. The substitution was made after the 2010 oil spill in the Gulf of Mexico rendered a unit about the idyllic nature of the Gulf unusable. The three science clusters were intact clusters used in PISA 2009.

The cluster rotation design for the standard booklets in the Main Survey corresponds to designs used in previous PISA surveys and is shown in Figure 2.1.

This is a balanced incomplete block design. Each cluster (and therefore each test item) appears in four of the four-cluster test booklets, once in each of the four possible positions within a booklet, and each pair of clusters appears in one (and only one) booklet. An additional feature of the PISA 2012 test design is that one booklet (booklet 12) is a complete link, being identical to a booklet administered in PISA 2009.

Each sampled student was randomly assigned to one of the thirteen booklets administered in each country, which meant each student undertook two hours of testing. Students were allowed a short break after one hour, typically of five minutes duration. Matters such as these on the administration of test sessions are described in more detail in Chapter 6.

In addition to the thirteen two-hour booklets, a special one-hour booklet, referred to as the UH booklet (Une Heure booklet), was prepared for use in schools catering for students with special needs. The UH booklet contained about half

■ Figure 2.1 ■

Cluster rotation design used to form standard test booklets for PISA 2012

Booklet ID	Cluster			
B1	PM5	PS3	PM6A	PS2
B2	PS3	PR3	PM7A	PR2
B3	PR3	PM6A	PS1	PM3
B4	PM6A	PM7A	PR1	PM4
B5	PM7A	PS1	PM1	PM5
B6	PM1	PM2	PR2	PM6A
B7	PM2	PS2	PM3	PM7A
B8	PS2	PR2	PM4	PS1
B9	PR2	PM3	PM5	PR1
B10	PM3	PM4	PS3	PM1
B11	PM4	PM5	PR3	PM2
B12	PS1	PR1	PM2	PS3
B13	PR1	PM1	PS2	PR3

as many items as the other booklets, with about 50% of the items being mathematics items (cluster PMUH), 25% reading (cluster PRUH) and 25% science (cluster PSUH). The items were selected from the Main Survey items taking into account their suitability for students with special educational needs, using criteria established in the lead-up to the PISA 2003 survey through consultation with the OECD Working Group on students with special educational needs.

In PISA 2012, as in PISA 2009, some countries were offered the option of administering an easier set of booklets whilst still providing an assessment that would generate results that are fully comparable to those from every other PISA participant, leading to an expanded booklet design incorporating material for both the standard PISA implementation and an implementation using the easier booklets. The offer was made to countries that had achieved a mean scale score in reading of 450 or less in PISA 2009, and to new countries that were expected – judging by their results on the PISA 2012 Field Trial conducted in 2011 – to gain a mean result at a similar level. The purpose of this strategy was to obtain better descriptive information about what students at the lower end of the ability spectrum know, understand and can do. A further reason for including easier items was to make the experience of the test more satisfying for individual students with very low levels of proficiency in mathematics. For countries that selected the easier set of booklets two of the standard mathematics clusters (PM6A and PM7A) were replaced with two easier mathematics clusters (PM6B and PM7B). Apart from level of difficulty, the sets of items in the standard and easier clusters were matched in terms of major framework characteristics to ensure that whichever set of items were taken in a particular country, the framework specifications were met. The other eleven clusters (five clusters of mathematics items, three clusters of reading items and three clusters of science items) were administered in all countries.

Although only two of the clusters differed for standard and easier administration, the cluster rotation in the booklets (where each cluster appears four times, once in each of the possible positions in the four-cluster booklets) means that more than half of the booklets are affected by the existence of these alternatives. Countries administering the standard set of booklets implemented booklets 1 to 13. Countries administering the easier set of booklets implemented booklets 8 to 13 and booklets 21 to 27, as shown in Figure 2.2 in the full test design used in the paper-based component of the Main Survey including the optional components. The only difference between the two sets of booklets was that for some countries, booklets 1 to 7 (those containing clusters PM6A and PM7A) were replaced with booklets 21 to 27 (with the easier clusters PM6B and PM7B as substitutes).

In PISA 2012, an assessment of financial literacy was offered as an international option. Countries participating in this option administered an additional four booklets, each containing the two clusters of financial literacy items (denoted PF1 and PF2) as well as one cluster of mathematics material (cluster PM5) and one cluster of reading material (PR2). As with the core domains, a special one-hour booklet, referred to as the FLUH booklet (Financial Literacy Une Heure booklet), was prepared for use in schools catering for students with special needs. This booklet consisted of one cluster of financial literacy material (denoted PFUH), and one cluster of mathematics material (denoted PMUH). The items were selected from the Main Survey items taking into account their suitability for students with special educational needs. Countries administering the financial literacy assessment implemented booklets 71-74 (in addition to booklets 1-13 if administering the standard booklets, or 8-13 and 21-27 if administering the easier set of booklets).

■ Figure 2.2 ■

Cluster rotation design used to form all test booklets for PISA 2012

Booklet ID	Cluster				Standard booklet set	Easier booklet set
B1	PM5	PS3	PM6A	PS2	Y	
B2	PS3	PR3	PM7A	PR2	Y	
B3	PR3	PM6A	PS1	PM3	Y	
B4	PM6A	PM7A	PR1	PM4	Y	
B5	PM7A	PS1	PM1	PM5	Y	
B6	PM1	PM2	PR2	PM6A	Y	
B7	PM2	PS2	PM3	PM7A	Y	
B8	PS2	PR2	PM4	PS1	Y	Y
B9	PR2	PM3	PM5	PR1	Y	Y
B10	PM3	PM4	PS3	PM1	Y	Y
B11	PM4	PM5	PR3	PM2	Y	Y
B12	PS1	PR1	PM2	PS3	Y	Y
B13	PR1	PM1	PS2	PR3	Y	Y
B20 (UH)	PMUH	PRUH/PSUH				
B21	PM5	PS3	PM6B	PS2		Y
B22	PS3	PR3	PM7B	PR2		Y
B23	PR3	PM6B	PS1	PM3		Y
B24	PM6B	PM7B	PR1	PM4		Y
B25	PM7B	PS1	PM1	PM5		Y
B26	PM1	PM2	PR2	PM6B		Y
B27	PM2	PS2	PM3	PM7B		Y
B70 (FLUH)	PFUH	PMUH				
B71	PF1	PF2	PM5	PR2		
B72	PF2	PF1	PR2	PM5		
B73	PM5	PR2	PF1	PF2		
B74	PR2	PM5	PF2	PF1		

As was mentioned earlier, material used to populate the design-included item clusters that originated in earlier PISA surveys, included again here to facilitate the linking of ability estimates across survey administrations, as well as new mathematics material needed to support the expansion of mathematics to "major domain" status for the PISA 2012 administration.

Computer-based assessment design: problem solving, mathematics, reading

For PISA 2012, a computer-based assessment of problem solving was included as part of the core assessment, which was taken up by about two-thirds of participating countries. Whilst the PISA Governing Board had wished to introduce the problem solving component for the PISA 2012 survey, a number of countries for a variety of technical and other reasons were not able to meet this wish. Nevertheless, problem solving continued to be referred to as a core component of the assessment.

In addition, countries were offered assessments of computer-based mathematics (CBAM) and reading in a digital environment (DRA). The latter two were offered together, in an assessment of computer-based literacies (CBAL). Countries implementing any part of the assessment on computer would either administer the assessment of problem solving only, or, assessments of all three of problem solving, CBAM, and DRA. They could not choose to administer CBAL while opting out of the assessment of problem solving.

The Main Survey items for the problem solving assessment were to populate four item clusters with each cluster representing 20 minutes of test time. For the countries administering the problem solving assessment as their only computer-based component, the test design specified that items would be presented to students in eight test forms, with each form being composed of two clusters according to the rotation design shown in Figure 2.3. The labels CP1 to CP4 denote the four computer-based problem solving clusters.

Each sampled student was randomly assigned one of the eight forms, which meant each student undertook 40 minutes of testing.

■ Figure 2.3 ■

Main Survey test design for countries participating in problem solving only

Form	Cluster 1	Cluster 2
31	CP1	CP2
32	CP2	CP3
33	CP3	CP4
34	CP4	CP1
35	CP2	CP1
36	CP3	CP2
37	CP4	CP3
38	CP1	CP4

Main Survey items for the CBAM and DRA were to populate four and two item clusters respectively, with each cluster representing 20 minutes of test time. For the countries administering the problem solving assessment together with the CBAL, the design specified that items would be presented to students in 24 test forms, with each form being composed of two clusters according to the rotation design shown in Figure 2.4. The labels CM1 to CM4 denote the four computer-based mathematics clusters, and CR1 and CR2 denote the two digital reading clusters.

Each sampled student was randomly assigned one of the 24 forms, which meant each student undertook 40 minutes of testing.

■ Figure 2.4 ■

Main Survey test design for countries participating in problem solving and CBAL

Form	Cluster 1	Cluster 2
41	CP1	CP2
42	CR1	CR2
43	CM3	CM4
44	CP3	CR1
45	CR2	CM2
46	CM1	CP4
47	CR2	CR1
48	CM2	CM1
49	CP3	CP4
50	CM4	CR2
51	CP1	CM3
52	CR1	CP2
53	CM1	CM3
54	CP4	CP1
55	CR1	CR2
56	CP2	CM4
57	CR2	CP3
58	CM2	CR1
59	CP2	CP3
60	CM4	CM2
61	CR2	CR1
62	CM3	CP1
63	CR1	CM1
64	CP4	CR2

Domain definitions, and item design: The 2012 assessment frameworks

The material needed to fulfil the design requirements had to satisfy the domain definitions and specifications within the relevant assessment framework. For each PISA subject domain, an assessment framework is produced to guide the PISA assessments in accordance with the policy requirements of the PISA Governing Board. The framework defines the domain, describes the scope of the assessment, specifies the structure of the test – including item format and the preferred distribution of items according to important framework variables – and outlines the possibilities for reporting results.

The PISA domain frameworks are conceived as evolving documents that will be adapted over time to integrate developments in theory and practice. Since a framework for PISA mathematical literacy had been partially developed

for the first PISA administration in 2000, and more fully articulated for PISA 2003 when mathematics was the major test domain for the first time, the PISA 2012 work began with a review of the existing framework at the initial meeting of the Mathematics Expert Group (MEG) in October 2009. That review and subsequent development work was carried out jointly by ACER and Achieve, the organisations appointed by the PISA Governing Board to jointly revise the mathematics framework for PISA 2012, in accordance with a development plan and timeline adopted by the PGB at its November 2009 meeting. Work on mathematics framework development commenced in October 2009 and continued through to adoption of the framework by the PISA Governing Board in November 2010.

A preparatory step in this development process was a survey of mathematical content standards applying in a range of relatively high-performing OECD countries, carried out by Achieve. Countries in that analysis included Australia, Belgium, Canada (Alberta), Finland, Ireland, Japan, Korea, New Zealand, and the United Kingdom. Achieve also analysed the previous frameworks and co-ordinated an extensive consultation process on the revised framework with experts from a range of countries as the PISA 2012 framework was under development. That consultation included consideration of responses to a detailed survey instrument, with responses from over 80 individuals (largely mathematicians and mathematics educators) from 34 countries participating in PISA. Several changes were proposed to the framework: *(i)* a revised definition of mathematical literacy was proposed and successively refined; *(ii)* the ways in which mathematical content was conceptualised and described underwent considerable revision over several drafts; *(iii)* the definition and description of mathematical processes were very substantially changed, resulting in a configuration of processes that would underpin a new set of reporting dimensions for PISA mathematics outcomes; and *(iv)* the contexts within which opportunities for students to express their levels of mathematical literacy would be provided were also reviewed and revised. Extension of the framework to incorporate a computer-based assessment option was developed, and a set of background variables that would be of particular interest was identified for mention in the framework. Revised framework drafts were presented in 2010 to successive meetings of the PGB which adopted a final version in 2011. An external validation of the item pool was implemented by Achieve to support the PGB's consideration of the items proposed to be used in the PISA 2012 survey instruments. Achieve engaged a team of mathematics experts to carefully review the items, and provided an independent external judgement about the fit of each item to the new framework.

The reading and science frameworks were unchanged in PISA 2012. However, new frameworks for two components of the PISA 2012 survey, the computer-based assessment of problem solving, and the assessment of financial literacy, were developed by ACER and its collaborators so that they could be adopted and published as part of the consolidated framework publication for PISA 2012.

Computer delivery was fundamental to the conception of problem solving in PISA 2012. It enabled *interactive* problems – problems in which exploration is required to uncover undisclosed information (Ramalingam et al., 2014) – to be included in a large-scale international assessment for the first time. In developing these problems, the emphasis was on everyday problem situations that often arise when interacting with an unfamiliar device (such as a ticket vending machine, air-conditioning system, or mobile phone) for the first time. Some of these devices, such as vending machines, were modelled as finite state machines (Buchner and Funke, 1993; Funke, 2001), that is, systems with a finite number of states, input signals and output signals. The system's next state is determined by its current state and the specific input signal selected by the user. Other problem situations, such as controlling an air conditioner, involved manipulating input and output variables that are related in some causal way. These situations were implemented as "MicroDYN" units (Greiff et al., 2013; Wüstenberg et al., 2012).

A particular challenge in task development for the PISA 2012 problem-solving assessment arose from the requirement to construct problems that did not need any particular domain-based knowledge for their solution, and with which students were not already familiar. This was intended to ensure that the focus was on measuring the cognitive processes involved in problem solving in a way more or less uncontaminated by the specific domain-based knowledge students had previously acquired through their other studies. This approach constitutes a major difference from the assessment of the other core domains in PISA (reading, mathematics and science), in which the assessments are constructed so that expert knowledge in the domain is required, indeed forms a main target of the assessment. For the assessment of problem solving, wherever possible low-verbal and non-verbal information was used in describing problems, hence minimising potential dependence on reading literacy skills, and only a basic level of mathematical and scientific knowledge was involved. In reality, ensuring that problems are equally unfamiliar to students is impossible at the individual level, but could perhaps be achieved across countries by presenting a variety of contexts so that no one group was consistently advantaged or disadvantaged in this way.

The assessment of financial literacy was an international option for PISA 2012. Development of the framework was overseen by a group that represented the expertise of the OECD Financial and Enterprise Affairs Directorate and the international experts who had been advising the Directorate in its efforts to promote financial education around the world.

The assessment framework drew heavily on work of the OECD-sponsored International Network on Financial Education (INFE), established in 2008, as well as on that of individual researchers at the national level. Like other PISA literacy domains, the financial literacy assessment framework set out ways of measuring the proficiency of 15-year-olds in demonstrating and applying knowledge and skills, while recognising that certain limitations had to be taken seriously given the enormous variation among OECD countries in the legislative, regulatory and practical approaches taken to financial matters. Key concepts to be included were the **content** of financial literacy (identified as *money and transactions*, *planning and managing finances*, *risk and reward* and *financial landscape*) and essential **processes** (*identify financial information*, *analyse information in a financial context*, *evaluate financial issues* and *apply financial knowledge and understanding*). The framework also identified four **contexts** in which the financial literacy of 15-year-olds should be demonstrated: *education and work*, *home and family*, *individual* and *societal*.

The items for the 2012 financial literacy assessment were developed by ACER and presented to the financial literacy expert group for feedback. It was the role of the expert group to ensure that the items developed matched the financial literacy framework that was being developed in parallel at the time. Advice was also sought from the expert group on whether the items were suitably aligned with the varied financial systems of the different countries taking part in the assessment. Due to time constraints, the various National Centres were unable to provide robust feedback on the items, but they were able to alert ACER to items that were inconsistent with their own financial systems and practices. Many of the items formed part of small units (consisting of between two and four items) whereas other items were stand-alone questions. In total, 81 items were included in the Field Trial and 40 items were included in the Main Survey. The items comprised simple vocabulary and no more than basic mathematics so as not to disadvantage those students with low reading and mathematics abilities.

One of the greatest challenges of item development for the financial literacy assessment was creating scenarios that applied equally to students from the different participating countries. For example, the financial consumer's relationship with credit cards and credit services varies widely between countries, and so the scenarios developed around credit had to be non-specific, ensuring that different countries' students were neither advantaged nor disadvantaged by the items. Similarly, items involving taxation had to be fairly generic to reflect the different taxation systems used in the different countries. Items involving value-based judgements were generally avoided as it was not considered sensible to use items to assess a student's attitude to saving and spending, noting that what may be a "sound" financial decision for the majority of people may not be the case for certain individuals in certain circumstances.

Another problem that had to be resolved with the expert group was the degree of financial knowledge and skills expected of a 15-year-old to "enable participation in economic life" (part of the framework's definition of financial literacy). Many financial concepts are beyond the first-hand experiences of the typical 15-year-old, with scenarios like pension contributions far off the student's radar. Financial scenarios such as shopping and saving up for a large purchase are commonplace activities throughout all countries but relying on such basic scenarios would limit the efficacy of the assessment. Some participating countries already had in place financial education courses for students but many others did not, and the lack of consistency among those existing financial education frameworks meant that much of the assessment framework was developed with fewer models to draw on than domains such as mathematics and science.

In 2012, the framework was prepared for publication along with an extensive set of example items. All five PISA 2012 cognitive frameworks were published in *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013).

Test development centres

Experience gained in the four previous PISA assessments showed the importance of using the development expertise of a diverse range of test centres to help achieve conceptually rigorous material that has the highest possible levels of cross-cultural and cross-national diversity. Accordingly, to prepare new mathematics and problem solving items for PISA 2012, ACER drew on the resources of nine test development centres in culturally-diverse and well-known institutions, namely ACER, the University of Melbourne (both in Australia), aSPe (University of Liege, Belgium), DIPF (Deutschen Institut für Internationale Pädagogische Forschung), IPN (Leibniz-Insitute for Science and Mathematics Education) and

Heidelberg University (all three in Germany), NIER (the National Institute for Educational Policy Research, Japan), CRP-HT (the Centre de Recherche Public – Henri Tudor, Luxembourg), ILS (the Department of Teacher Education and School Research, University of Oslo, Norway) and ETS (Education Testing Service, United States). For financial literacy, all new test development was undertaken at ACER.

ACER co-ordinated the distribution of material for development across these centres, and managed the co-operative development processes in which the item writers in each centre engaged. The test development teams were encouraged to conduct initial development of items, including cognitive laboratory activities, in their local language. Translation to the OECD official languages (English and French) took place after items had reached a well-formed state.

Scope, volume and constraints

PISA items are arranged in units based around a common stimulus. Many different types of stimulus are used including passages of text, tables, graphs and diagrams, often in combination. Each unit contains from one to five items assessing students' competencies and knowledge. A complete PISA unit consists of some stimulus material, one or more items (questions), and a guide to the coding of responses to each question. Each coding guide comprises a list of response categories (full, partial and no credit), each with its own scoring code, descriptions of the kinds of responses to be assigned each code, and sample responses for each response category.

For the paper-based assessment, 56 mathematics units comprising a total of 110[1] cognitive items were needed to provide approximately 270 minutes of testing time for mathematics in PISA 2012. The reading assessment consisted of 44 items (13 units), a subset of the 131 items used in 2009, representing 90 minutes of testing time. The science assessment consisted of 53 items (18 units), also representing 90 minutes of testing time. The science items were the same as those used in 2009. The optional assessment of financial literacy consisted of 29 units, comprising a total of 40 items, representing 60 minutes of testing time (see Annex A).

The 110 cognitive mathematics items used in the Main Survey included 36 items from the 2003 test that had also been used for linking in 2006 and 2009. The remaining 74 items were newly developed for PISA 2012. The 74 new items were selected from a pool of 172 newly-developed items that were tested in a Field Trial conducted in all countries in 2011, one year prior to the Main Survey. The 40 items comprising the financial literacy assessment were newly-developed for PISA 2012, and were selected from a pool of 75 items that were similarly tested in a Field Trial conducted in 2011 in countries participating in this international option. There was no new item development for reading or science, as the design requirements could be met with existing secure material.

The problem solving assessment comprised sixteen units, with a total of 42 items, representing 80 minutes of testing time in total. These items were selected from a pool of 79 newly-developed problem solving items that were tested in a Field Trial conducted in all participating countries in 2011, one year prior to the Main Survey. The instrument for the CBAM comprised 15 units, with a total of 41 items, representing 80 minutes of testing time in total. These items were selected from a pool of 86 newly-developed computer-based mathematics items that were tested in a Field Trial conducted in all participating countries in 2011, one year prior to the Main Survey. As well as the item format types referred to in relation to the paper-based assessment items, additional variants of the selected response format type were used with items that involved, for example, selection from a drop-down menu, use of "drag and drop" and use of "hot spots".

The instrument for the DRA assessment consisted of 19 items, based on 6 units, representing 40 minutes of testing time. The digital reading items were selected from the 29 items used in the DRA in PISA 2009.

In each of the computer-based assessments, units and items within units were delivered in a fixed order, or lockstep fashion. This meant that students were not able to return to an item or unit once they had moved to the next item/unit. Each time a student clicked the "Next" test navigation button, a dialog box displayed a warning that the student was about to move on to the next item and that it would not be possible to return to previous items. At this point students could either confirm that they wanted to move on or cancel the action and continue with the item they had been viewing.

The assessment items for problem solving and computer-based mathematics each make use of only one screen of stimulus material, but the stimulus used in the digital reading assessment comprises digital texts with the structures and features of websites, e-mails, blogs and so on. In the case of the DRA, then, lockstep delivery enabled test developers to specify the starting browser page for each item. This meant that all students began in the same place within the stimulus and, if they had previously navigated through a series of less relevant pages, did not have to spend time finding their way to an appropriate page to begin the item task.

Item formats employed with paper-based cognitive items were either selected response or constructed response. Selected response items were either standard multiple-choice with four (or in a small number of cases, five) responses from which students were required to select the best answer, or complex multiple-choice presenting several statements for each of which students were required to choose one of two or more possible responses (yes/no, true/false, correct/incorrect, etc.). Constructed response items were of two broad types. Constructed response manual items required limited manual input by trained coders at the stage of processing student responses. They required students to construct a numeric response within very limited constraints, or only required a word or short phrase as the answer, and coders later to assign each response to the predefined response categories. Constructed response expert items required the use of trained expert coders to interpret observed student responses and assign them to one of the defined response categories. These items required a response to be generated by the student, with a range of possible full-credit answers.

For the computer-based cognitive items, two additional item formats were employed. The first, constructed response auto-coded, included any item in which students constructed a non-text based response. This might be done, by, for example, highlighting segments of map to show an optimal route, or dragging and dropping an object from one point to another. As the name suggests, scoring rules were defined for such items so that they could be coded automatically. The other new response format was "selected response variations". These included any item in which the student selected a response that was not multiple-choice or complex multiple-choice. This item type included drop down menu items where either a) there was more than one drop down menu; b) there was more than one possible correct response; or c) where more than one choice could be made. For example, select the best two responses from the following list.

Pencils, erasers, rulers, and in some cases calculators, would be provided to students undertaking the PISA assessment. It was recommended that calculators be provided in countries where they were routinely used in the classroom. National Centres decided whether calculators should be provided for their students on the basis of standard national practice. No test items required a calculator, but some mathematics items involved solution steps for which the use of a calculator could be of assistance to students accustomed to their use.

Development timeline and processes

Planning for mathematics item development began in September 2009, with preparation of material for a two-day meeting of test developers from each test development centre, which was held in Offenbach on 19-21 October, 2009. The meeting had the following purposes:

- to become familiar with the issues under consideration by ACER and Achieve in revising the mathematics framework for PISA 2012, especially the implications of possible changes for test development;
- to discuss the requirements for item development, including item presentation and formats, use of templates and styles and cognitive laboratory procedures and timelines;
- to discuss factors that influence item difficulty, particularly in light of the intention to develop items at the extremes of the scale (a contractual requirement);
- to be briefed on detailed guidelines, based on experience from the first four PISA administrations, for avoiding potential translation and cultural problems when developing items; and
- to review sample items prepared for the meeting by each of the test development centres.

The meeting reviewed documentation prepared by ACER to guide all parts of the process for the development of cognitive items: the calling for submissions from participating countries, writing and reviewing items, carrying out cognitive laboratory activities and pilot tests of items and conducting an extensive Field Trial, producing final source versions of all items in both English and French, preparing coding guides and coder training material, and selecting and preparing items for the Main Survey, all in time to distribute material to PISA National Centres in each participating country well in advance of the commencement of the Main Survey in March 2012. The main phase of test development finished when the items were distributed for the Field Trial in December 2010. During this 15-month period, intensive work was carried out writing and reviewing items, and on various cognitive laboratory activities. The Field Trial for most countries took place between March and August 2011, after which items were selected for the Main Survey and distributed to countries in December 2011.

The material from which the new mathematics items were developed originated from three main sources. First, the National Centres from participating countries submitted a large number of items or ideas for items, some 500 in total including about 400 intended for paper-based delivery and about 50 intended for computer delivery. Material was

submitted by twenty different National Centres (Canada, Colombia, the Czech Republic, France, Greece, Israel, Italy, Korea, Macao-China, Mexico, the Netherlands, Portugal, Serbia, Shanghai-China, Singapore, Spain, Switzerland, Chinese Taipei, Uruguay and the United States). Second, the members of the Mathematics Expert Group and Consortium staff working with that group contributed a small pool of items, many of which were designed to expand the volume of relatively easy material available for selection. Third, the teams of professional item writers engaged by ACER to develop material provided a significant volume of original material, in addition to the development work those teams carried out to refine submitted material.

The development timeline for the problem solving items was similar to that for mathematics, although heavy involvement of test development centres outside ACER occurred at a slightly later point in the development process. The items for the PISA 2012 problem-solving assessment came from two sources: the PISA international Consortium and national submissions. After initial development work by the test development centres, the Problem-Solving Expert Group that developed the PISA 2012 framework reviewed materials to ensure that they reflected the defined construct of problem-solving competence. Small-scale cognitive laboratory activities were conducted, and the items were reviewed by National Centres and field tested.

First phase of development

Typically, the following steps were taken in the first phase of the development of mathematics items. A similar process, simplified and shortened in some cases, was followed in the other (minor) domains for which new item development was needed. The steps are described in a linear fashion, but in reality they were often negotiated in a cyclical fashion, with items going through the various steps more than once.

Initial preparation

At the early stages of test development, test developers in each of the Consortium test development centres found potential material and exchanged it with one or more other centres (in English translation if necessary) to ascertain whether colleagues agreed that it was worth developing further, or they worked with material that had originated in national item submissions that had been assigned to them for development. The material was formatted even at this early stage in a manner similar to that planned for the final presentation.

For material that was judged worth pursuing, test developers prepared units in both English and their native language in a standard format, including stimulus, several items (questions), and a proposed coding guide for each item. Items were then subjected to a series of cognitive laboratory activities: item panelling (also known as item shredding or cognitive walkthrough), cognitive interviews, and pilot or pre-trial testing (also known as cognitive comparison studies).

Local item panelling

Each unit first underwent extensive scrutiny at a meeting of members of the originating test development team. This stage of the cognitive laboratory process typically involved item writers in a vigorous analysis of all aspects of the items from the point of view of a student, and from the point of view of a coder.

Items were revised, often extensively, following item panelling. When substantial revisions were required, items went back to the panelling stage for further consideration.

Cognitive interviews

Many units were then prepared for individual students or small groups of students to attempt. For paper-based material a combination of think-aloud methods, individual interviews and group interviews was used with students to ascertain the thought processes typically employed as students attempted the items. For computer-based items, all cognitive interviews were conducted individually, using either audio-recording of responses or dual administration, with one researcher interacting with the student and a second researcher observing and recording navigation behaviour.

Items were revised, often extensively, following their use with individuals and small groups of students. This stage was particularly useful in clarifying the wording of questions, and gave information on likely student responses that was used in refining the response coding guides.

Local pilot testing

As the final step in the first phase of print item development for several of the items, sets of units were piloted with several classes of 15-year-olds. As well as providing statistical data on item functioning, including the relative difficulty of items,

this enabled real student responses derived under formal test conditions to be obtained, thereby enabling more detailed development of coding guides.

Pilot test data were used to inform further revision of items where necessary or sometimes to discard items altogether. Units that survived relatively unscathed were then formally submitted to the test development manager to undergo their second phase of development.

Second phase of development

The second phase of item development began with the review of each unit by at least one test development team that was not responsible for its initial development. Each unit was then included in at least one of a series of pilot studies with a substantial number of students of the appropriate age.

International item panelling

The feedback provided following the scrutiny of items by international colleagues often resulted in further improvements to the items. Of particular importance was feedback relating to the operation of items in different cultures and national contexts, which sometimes led to individual items or even whole units being discarded. Surviving units were considered ready for further pilot testing and for circulation to National Centres for review.

International pilot testing

For each pilot study, test booklets were formed from a number of units developed at different test development centres. These booklets were trial tested with several whole classes of students in several different schools. Field-testing of this kind mainly took place in schools in Australia because of translation and timeline constraints. Sometimes, multiple versions of items were trialled and the results were compared to ensure that the best alternative form was identified. Data from the pilot studies were analysed using standard item response techniques.

Many items were revised, usually in a minor fashion, following review of the results of pilot testing. If extensive revision was considered necessary, the item was either discarded or the revised version was again subject to panelling and piloting. One of the most important outputs of this pilot testing was the generation of many student responses to each constructed-response item. A selection of these responses was added to the coding guide for the item to further illustrate each response category and provide more guidance for coders.

National item submissions

An international comparative study should ideally draw items from as many participating countries as possible to ensure wide cultural and contextual diversity. A comprehensive set of guidelines, was developed to encourage and assist national submission of items. The document *Item Development for PISA 2012 and Item Submission Guidelines* was distributed to PISA 2012 National Project Managers in March 2010.

The guidelines described the scope of the item development task for PISA 2012, the arrangements for national submissions of items and provided sample items. In addition, the guidelines contained a detailed discussion of item requirements and an overview of the full item development process for PISA 2012.

To assist countries in submitting high quality and appropriate material, ACER conducted a one-day mathematics item development workshop for interested National Centres at the end of the first meeting of National Project Managers (NPMs) for PISA 2012, in March 2007. It was attended by individuals from most National Centres. The due date for national submission of items was 31 May 2010 for problem solving, and 1 June 2010 for mathematics, as late as possible given Field Trial preparation deadlines. Items could theoretically be submitted in any language, but in many cases the preliminary development work that occurred in country concluded with the preparation of an English language version prior to submission. Countries were urged to submit items as they were developed, rather than waiting until close to the submission deadline. It was emphasised that before items were submitted they should have been subject to some cognitive laboratory activities involving students, and revised accordingly. For mathematics, an item submission form was provided with the guidelines and a copy had to be completed for each unit, indicating the source of the material, any copyright issues, and the framework classifications of each item.

Approximately 450 items were submitted by PISA National Centres for consideration by the international contractor’s test development teams. These items came from about 20 different countries. Some submitted units had already undergone significant development work. Others were in a less developed state. All submitted material was initially reviewed by the

test development co-ordinator at ACER, to check for consistency with the framework, to identify material that was repetitive (for example, to identify instances where two different National Centres had submitted material that was very similar, or that was too similar to material already in development) or that may have been unsuitable for other reasons (such as being too ephemeral, or sensitive on cultural grounds). Where material was deemed suitable at this initial screening stage, it was assigned to one of the test development teams, after which the processes described earlier were applied.

National review of items

In July 2010, National Project Managers (NPMs) were given a set of item review guidelines to assist them in reviewing cognitive items and providing feedback, using an online review and feedback system that was developed by ACER for this purpose. Bundles of items were made available progressively through 2010 as item development proceeded, with item bundles being released in March, April, July, and two in August 2010. A central feature of those reviews was the requirement for national experts to rate items according to various aspects of their relevance to 15-year-olds, including whether they related to material included in the country's curriculum, their relevance in preparing students for life, how interesting they would appear to students and their authenticity as real applications of mathematics. Corresponding feedback categories were used for the other domains. NPMs were also asked to identify any cultural concerns or other problems with the items, such as likely translation or coding difficulties, and to give each item an overall rating for retention in the item pool. For items intended for computer delivery (CBAM and problem solving), feedback was also sought on the likely demands related specifically to general computer use and familiarity that would be essentially unrelated to the cognitive objectives of the items.

For each bundle, a series of reports was generated summarising the feedback from National Project Managers. The feedback frequently resulted in useful input to the international contractor's test development teams in its task of further revising the items. In particular, cultural issues related to the potential operation of items in different national contexts were highlighted and sometimes, as a result of this, items had to be discarded. Summaries of the ratings assigned to each item by the NPMs were used extensively in the selection of items for the Field Trial.

International item review

As well as the formal, structured process for national review of items, cognitive items were also considered in detail, as they were developed, at meetings of the PISA MEG that took place in 2010 and 2011.

In addition, as mentioned earlier Achieve conducted an independent external validation study in relation to the mathematics items selected for use in the Field Trial, to assess the extent to which they were a proper reflection of the objectives and constraints specified in the mathematics framework. The conclusion in the report of the validation study was:

> *"... that the items represent the framework well, and cover the mathematics expected of 15-year-olds at an appropriate breadth and depth. Also, assuming the selection of operational items from this field test pool addresses concerns voiced by the external validation panel, they agreed that PISA 2012 will assess the construct of mathematical literacy as defined in the framework."*

Preparation of dual (English and French) source versions

Both English and French source versions of all paper-based test instruments were developed and distributed to countries as a basis for local adaptation and translation into national versions. An item-tracking database, with web interface, was used by both test developers and Consortium translators to access items. This ensured accurate tracking of the English language versions and the parallel tracking of French translation versions, ensuring synchronisation of the two source versions.

Part of the translation process involved a technical review by French subject experts, who were able to identify issues with the English source version related to content and expression that needed to be addressed immediately, and that might be of significance later when items would be translated into other languages. Many revisions were made to items as a result of the translation and technical review process, affecting both the English and French source versions. This parallel development of the two source versions assisted in ensuring that items were as culturally neutral as possible, identified instances of wording that could be modified to simplify translation into other languages, and indicated where additional translation notes were needed to ensure the required accuracy in translating items to other languages.

Field testing

The PISA Field Trial was carried out in all countries with the implementation occurring for the majority of countries in the first half of 2011. An average of over 200 student responses to each item was collected in each country. During the Field Trial, the Consortium set up a coder query service. Countries were encouraged to send queries to the service so that a common adjudication process was consistently applied to all coders' questions about constructed-response items. Between July and November 2011, the test development centres, the mathematics, problem solving and financial literacy expert groups and National Centres reviewed the Field Trial data to support the identification of a proposed selection of Field Trial items for the Main Survey.

Field Trial item selection

A total of a 474 mathematics items (344 paper-based and 130 computer-based) were circulated to National Centres for review from early March to late August 2010. Seventy-four of those (65 paper-based and 9 computer-based) had originated from national submissions.

From that pool of 474 items, 172 paper-based items were selected to supplement the pre-existing 36 link items, and 86 computer-based items were selected, for inclusion in the Field Trial. The selection of those items took into account a number of factors: the rating of items by national experts (their priority for inclusion) as part of their review of the item bundles, other item feedback from National Centres bearing on item quality and acceptability, the preferences of expert group members based largely on the fit of items to the objectives and definitions of the framework, data derived from cognitive laboratories and small-scale pilot activities including data on the expected difficulty of items, and the need to balance the selection against the framework's test specification.

A similar selection process occurred for the problem solving items, where 79 items were selected for inclusion in the Field Trial; and likewise for financial literacy, where 75 items were selected.

For the paper-based reading and science components of the Field Trial, material was used in intact clusters from previous PISA administrations (with the exception of one reading unit replaced as mentioned earlier); and likewise for the digital reading component, which used intact material from the 2009 digital reading assessment.

Field Trial design

Paper-based assessment

The Field Trial design for the paper-based assessment comprised 17 clusters of mathematics items (denoted PM1 to PM17), 3 clusters of reading items (PR1 to PR3) and 3 clusters of science items (PS1 to PS3).

Clusters PM1, PM2 and PM3 were intact clusters that had been used in PISA 2003, 2006, and 2009 comprising 36 link items (in 25 units). The 172 new mathematics items (from 62 units) were allocated to 14 clusters, PM4 to PM17.

PR1 and PR2 were two intact reading clusters from PISA 2006 and PR3 was an almost intact cluster from 2006 but with one three-item unit inserted in place of material that had to be replaced. These three clusters comprised 44 items (13 units). PS1, PS2 and PS3 were 3 intact science clusters comprising 53 items (18 units) selected from the 2006 survey.

Material for the optional financial literacy component comprised 75 items placed in four clusters. In addition, the Field Trial design included a one-hour test booklet comprising one mathematics cluster, a half cluster of reading material and a half cluster of science material, for special educational needs students. Items in these clusters were selected taking into account their suitability for students with special educational needs.

Ten regular two-hour booklets, each comprising four clusters, were administered in the Field Trial. Each cluster was designed to take up 30 minutes of testing time, thus making up booklets with two hours' worth of testing time. New mathematics clusters appeared once in the first half of a booklet and once in the second half, in booklets 1 to 8, and were administered in all participating countries. The mathematics, reading and science link material appeared in booklets 9 and 10; these booklets were administered only in countries participating in PISA for the first time in 2012. Figure 2.5 shows the Field Trial design for the paper-based assessment.

Two one-hour booklets were administered in the Field Trial to support the testing of sampling and operational procedures in schools having students with special educational needs, one for students in the regular sample (labelled BUH in Figure 2.5), and one for students in the sample for the financial literacy international option (labelled BFUH).

The booklets used were identical to those that had been used in the PISA 2009 test booklet rotation design. Booklet BUH comprised a reading cluster labelled in Figure 2.5 as PRUH, and two half clusters (one for each of mathematics and science) labelled as PMUH and PSUH. Exactly the same clusters were used in the BFUH booklet.

■ Figure 2.5 ■

Allocation of item clusters to test booklets for Field Trial

Booklet ID	Cluster				Booklet set for:
B1	PM4	PM12	PM13	PM6	All participating countries
B2	PM5	PM13	PM14	PM7	All participating countries
B3	PM6	PM14	PM15	PM8	All participating countries
B4	PM7	PM15	PM16	PM9	All participating countries
B5	PM8	PM16	PM17	PM10	All participating countries
B6	PM9	PM17	PM1	PM11	All participating countries
B7	PM10	PM1	PM2	PM4	All participating countries
B8	PM11	PM2	PM12	PM5	All participating countries
B9	PM3	PS1	PS2	PS3	Only new countries
B10	PR1	PR2	PR3	PM3	Only new countries
BFL1	PF1	PF2	PF3	PF4	
BFL2	PF4	PF3	PF2	PF1	
BUH	PRUH	PMUH/PSUH			
BFUH	PRUH	PMUH/PSUH			

Computer-based assessment

The 86 computer-based mathematics items were arranged in eight clusters each designed to occupy 20 minutes of test time, and these were administered in pairs in eight test forms, hence each form occupied 40 test minutes.

The 79 Field Trial items for problem solving were also arranged in eight twenty-minute clusters, and these were also administered in pairs in eight test forms.

Two twenty-minute clusters of computer-based reading material were formed from the 18 items, and delivered in two test forms.

Dispatch of Field Trial instruments

Field Trial instruments were dispatched to PISA National Centres in stages during the period from late October to December 2010 as they reached their final form.

Final versions of material for computer delivery were released in the online translation management system in October 2010. Final English and French paper-based source versions of the new mathematics Field Trial units were distributed to National Centres in two batches, the first in November 2010 (along with the financial literacy material), and the second in early December 2010. All consolidated final source versions of booklets (in English and French) and forms (in English) were distributed on 22 December 2010. All material could also be downloaded from the PISA website from the time of dispatch.

As material became available, National Centres commenced the process of preparing national versions of all units, clusters and booklets. All items went through an extremely rigorous process of adaptation, translation and external verification in each country to ensure that the final test forms used were equivalent. That process and its outcomes are described in Chapter 5.

Field Trial coder training

Following final selection and dispatch of items to be included in the Field Trial, various documents and materials were prepared to assist in the training of personnel who would lead the coding of student responses in each PISA country. International coder training sessions for mathematics, reading, science, problem solving and financial literacy were conducted in February 2011. For the paper-based assessments, consolidated coding guides were prepared, in both English and French, containing all those items that required manual coding. The guides emphasised that coders were to code rather than score responses. That is, the guides defined different kinds of possible responses to each item, which did not all necessarily receive different scores. A separate training workshop document in English only was also produced for each paper-based domain. These workshop documents contained additional student responses to the items that required manual coding, and were used for practice coding and discussion at the coder training sessions. Corresponding training

material was also prepared for the computer-based components. Coding of response to computer-based items was carried out in an online coding system developed for the purpose. Explanatory material guided the use of the system as well as showing how manually coded items should be treated, for each of problem solving, mathematics and reading.

Countries sent representatives to the training sessions. Open discussion of how the workshop examples should be coded was encouraged and showed the need to introduce a small number of amendments to coding guides. These amendments were incorporated in a final dispatch of coding guides and training materials in March 2011. Following the international training sessions, National Centres conducted their own coder training activities using their verified translations of the consolidated coding guides. The support materials for coding prepared by the Consortium included a coder recruitment kit to assist National Centres in recruiting people with suitable qualifications to fill the role of expert coder.

Field Trial coder queries

The Consortium provided a coder query service to support the coding of constructed-response items in each country. When there was any uncertainty as to the code most appropriate to a particular observed item response, National Centres were able to submit queries by e-mail to the query service, and these were immediately directed to the relevant Consortium expert. Considered responses were quickly prepared, ensuring greater consistency in the coding of responses to items.

The queries with the Consortium's responses were published periodically on the PISA website. The queries report was regularly updated as new queries were received and processed. This meant that all national coding centres had prompt access to an additional source of advice about responses that had been found problematic in some sense. Coding supervisors in all countries found this to be a particularly useful resource though there was considerable variation in the number of queries that they submitted. Over successive PISA administrations, the accumulated coder queries have provided an excellent source of additional examples for the coding guides and training materials.

Field Trial outcomes

Extensive analyses were conducted on the Field Trial cognitive item response data, and included the standard *ACER ConQuest* item analysis (item fit, item discrimination, item difficulty, distractor analysis, mean ability and point-biserial correlations by coding category, item omission rates, and so on), as well as analyses of gender-by-item interactions and item-by-country interactions. In reviewing those statistics, for example, response categories needed to be well ordered according to the average abilities of students giving each response; the point-biserial correlation for the key category should be positive, and for the other categories much smaller or negative; the fit of items should be near to 1. These data would be vital information to be used in the selection of items for use later in the Main Survey. In addition, the coding of partial credit items was reviewed. In some cases, the collapsing of categories was recommended.

Consortium analysts routinely examined all items for evidence of Differential Item Functioning (DIF), whereby different subsets of the assessed population (for example, different gender groups, country or language groups) when matched for ability, found the items differentially difficult. Any such cases were carefully examined to determine whether wording, translation or other factors in the presentation of the item may have contributed, and if so whether the issue could be resolved through some minor adjustment of the item, or could not easily be resolved in which case the item was set aside as unsuitable for selection in the Main Survey item pool.

The parts of each complex multiple-choice item were also analysed separately and this led to some parts being dropped though the item itself was retained.

National review of Field Trial items

A further round of national item review was carried out in the online item review system, this time informed by the experience at National Centres of how the items had worked in the Field Trial in each country. A document, *Item Review Guidelines*[2] was produced to assist national experts to focus on the most important features of possible concern. In addition, NPMs were asked to assign a rating from 1 (low) to 5 (high) to each item to indicate its priority for inclusion in the Main Survey. A high proportion of participating countries completed this review of the Field Trial items.

A comprehensive Field Trial review report also was prepared by all NPMs, for both the paper-based and computer-based assessments. These reports included a further opportunity to comment on particular strengths and weaknesses of individual items identified during the translation and verification process and during the coding of student responses.

MAIN SURVEY PREPARATION, IMPLEMENTATION AND FOLLOW-UP

Main Survey item selection

The expert groups for mathematics, problem solving and financial literacy met in Melbourne in September 2011 to review all available material and recommend which items should be included in the Main Survey instruments.

The expert groups considered the pool of items (new items, and in the case of mathematics, items to be used to link 2012 outcomes to those of previous PISA administrations) that had been tested in the recent Field Trial and had performed adequately from a technical measurement perspective on the basis of the item statistics referred to in the previous section, and using criteria established in previous PISA survey analyses that are also referred to in Chapters 9 and 12 of this volume. The available items were evaluated by the expert groups in terms of their substantive quality, fit to framework, range of difficulty, National Centre feedback, and durability.

The selection of items to be proposed for inclusion in the Main Survey instruments had to satisfy the following conditions:

- the psychometric properties of all selected items had to be satisfactory (according to the criteria referred to above and in Chapters 9 and 12);
- items that generated coding problems in the Field Trial had to be avoided unless those problems could be properly addressed through modifications to the coding guides;
- items given high priority ratings by National Centres were to be preferred, and items with lower ratings were to be avoided;
- the major framework categories had to be populated as specified in the relevant framework; and
- there had to be an appropriate distribution of item difficulties, broad enough to generate useful measurement data at both extremes of the anticipated ability distribution of sampled students across all participating countries.

Recommended selections of items for mathematics (both the paper-based and computer-based components), problem solving and financial literacy were presented to a meeting of National Project Managers in October 2011 for their review and endorsement. Final recommendations were presented to the PISA Governing Board at its meeting in Israel in October 2011 for endorsement.

Characteristics of the mathematics item set used in the Field Trial, and the set used in the Main Survey, for both the paper-based and computer-based components, are summarised in Figure 2.6 showing the distribution of items in relation to the various categories specified in the framework.

■ Figure 2.6 ■

Mathematics item counts (Field Trial and Main Survey) by framework category

Framework category		Link items	New items			
			Paper-based		Computer-based	
			Field Trial	Main Survey	Field Trial	Main Survey
Content	Change and relationships	9	46	20	22	11
	Quantity	11	44	18	26	9
	Space and shape	9	42	18	19	12
	Uncertainty and data	7	40	18	19	9
Process	Formulate	11	42	22	16	9
	Employ	14	76	35	41	22
	Interpret	11	54	17	29	10
Context	Occupational	3	40	21	23	9
	Personal	5	50	16	22	13
	Public	14	42	15	17	11
	Scientific	14	40	22	24	8
Format type	Simple multiple choice	10	47	22	19	8
	Complex multiple choice	7	19	6	12	4
	Constructed response (automatic)				42	22
	Constructed response (expert)	8	60	23	9	4
	Constructed response (manual)	11	46	23		
	Constructed response (variations)				4	8

The item counts for mathematics, problem solving, reading, science and financial literacy (in each of the Field Trial and Main Survey) are presented in Figure 2.7.

■ Figure 2.7 ■

Item counts (Field Trial and Main Survey) by domain and delivery mode

Domain	Field Trial	Main Survey
Mathematics (paper-based)	208	110
Mathematics (computer-based)	86	41
Problem solving (computer-based)	79	42
Reading (paper-based)	44	44
Reading (computer-based)	18	18
Science (paper-based)	53	53
Financial literacy (paper-based)	75	40

Dispatch of Main Survey instruments

After finalising the Main Survey item selection, final forms of all selected items were prepared. This involved minor revisions to items and coding guides based on detailed information from the Field Trial, and the addition of further sample student responses to the coding guides.

French translations of all selected items were then updated. For the paper-based material, clusters of items were formatted, and booklets were formed in accordance with the Main Survey rotation design shown previously in Figure 2.2. For the computer-based material, the release of units included both digital versions of the units, and paper-based coding guides. English and French versions of all material were made available to National Centres in several dispatches, on 2 September (link clusters), 24 November and 5 December (new paper and computer-based units) and 20 December 2011 (new clusters and all booklets).

Main Survey coder training

Consolidated coding guides were prepared, in both English and French, containing all the items that required manual coding. These were dispatched to National Centres on 25 January 2012. In addition, the training materials prepared for Field Trial coder training were revised with the addition of student responses selected from the Field Trial coder query service.

International coder training sessions for reading, mathematics and science were conducted in Salzburg, Austria in February 2012. As had been the case for the Field Trial, it was apparent at the training meeting that a small number of clarifications were needed to make the coding guides and training materials as clear as possible. Revised coding guides and coder training material for both paper-based assessments and computer-based assessments were prepared and dispatched early in March 2012.

Main Survey coder query service

The coder query service operated for the Main Survey across all test domains. Any student responses that were found to be difficult to code by coders in National Centres could be referred to the Consortium for advice. The Consortium was thereby able to provide consistent coding advice across countries. Reports of queries and the Consortium responses were made available to all National Centres via the Consortium website, and were regularly updated as new queries were received.

Review of Main Survey item analyses

Upon reception of data from the Main Survey testing, extensive analysis of item responses was carried out to identify any items that were not capable of generating useful student achievement data. Such items could be removed from the international dataset, or in some cases from particular national datasets where an isolated problem occurred. One mathematics item was removed from the international data set as a result of this analysis. Further details on the outcomes of the analysis of Main Survey item data are provided in Chapter 12.

Released items

Several PISA items were released into the public domain at the time of publication of the PISA 2012 results, to illustrate the kinds of items used in the PISA assessment. Two intact clusters from the paper-based mathematics component of the

Main Survey, comprising 26 items, were released, along with a further 30 paper-based items that had been used in the Field Trial but were not selected for inclusion in the Main Survey item set. A further 11 mathematics items were released, from a cluster that had been used in the PISA 2006 administration but had subsequently been held in reserve. The items are available for download from the PISA website: *http://www.oecd.org/pisa/pisaproducts/*.

In addition, ten items from three units used in the computer-based mathematics component were released to supplement four units that had been put in the public domain prior to the assessment, along with four units of problem solving material to supplement the two that had been released earlier. Three additional reading units (to supplement the seven sample items previously posted) were added to the public website set up for this purpose. All of these computer-based items can be seen at *www.oecd.org/pisa*.

Some of these released paper-based items, including ten individual financial literacy items from the Field Trial that were not included in the Main Survey, were included in the publication *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013), and some were used for illustrative purposes in the OECD international report of the *PISA 2012 Results* (OECD, 2014).

No new reading or science material was released after the 2012 survey administration.

Notes

1. One of those items was deleted internationally as a result of errors detected in the coding of responses.

2. Technical reference documents are available on the OECD PISA website: *www.oecd.org/pisa*.

References

Buchner, A. and **J. Funke** (1993), "Finite-state Automata: Dynamic Task Environments in Problem-solving Research", *The Quarterly Journal of Experimental Psychology*, Vol. 46A, pp. 83-118.

Funke, J. (2001), "Dynamic systems as tools for analysing human judgement", *Thinking and Reasoning*, Vol. 7, pp. 69-79.

Greiff, S., D. V. Holt, S. Wüstenberg, F. Goldhammer, and **J. Funke** (2013) "Computer-based assessment of complex problem solving: Concept, implementation, and application", *Educational Technology Research & Development*, Vol 61, pp. 407-421.

Postlethwaite and **A. Grisay** (eds.), *A view inside primary schools: A World Education Indicators (WEI) cross-national study*, Unesco Institute of Statistics, Montreal, pp. 175-207.

OECD (2014), *PISA 2012 Results: What Students Know and Can Do Student Performance in Mathematics, Reading and Science (Volume I Revised edition, February 2014),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Sovling and Financial Literacy*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264190511-en

Ramalingam, D., B. McCrae and **R. Philpot** (2014), "The PISA Assessment of Problem Solving", Chapter 6, in Csapó, B. and J. Funke (eds.), *The Nature of Problem Solving*, OECD Publishing, Paris.

Wüstenberg, S., S. Greiff and **J. Funke** (2012), "Complex problem solving – More than reasoning?", *Intelligence*, Vol. 40, pp. 1-14.

3

Context Questionnaire Development

INTRODUCTION

The goal of the PISA background instruments is to gather data that can help policy makers and educators understand why and how students achieve certain levels of performance. PISA questionnaires must cover the most important antecedents and processes of student learning at the individual, school, and system level. The questionnaires also allow the collection of non-cognitive student performance outcomes such as student attitudes, interests, motivations, and beliefs.

At the same time, with the programme undertaking its fifth assessment, a number of points regarding the PISA context questionnaires required attention, including:

- Developing a sustainable framework for the context questionnaires that would ensure the monitoring of essential contextual characteristics over time while at the same time enabling new topics to be incorporated.
- Addressing questions regarding the cross-cultural comparability of measures in the context questionnaires.
- Transitioning the context questionnaires from paper administration to online administration mode.
- Updating the coding of parental occupation according to the 2008 International Standard Classification of Occupations (ISCO-08) from its 1988 version to the 2008 version.

In addition, the Consortium set itself the challenge of two further innovations in PISA 2012:

- the expanded measurement of Opportunity to Learn (OTL); and
- the rotation of the Student Context Questionnaire.

This chapter provides an overview of the questionnaires and their development process, while Chapter 16 describes questionnaire index construction and Chapter 17 describes the research that was undertaken during questionnaire construction and validation.

A SUSTAINABLE FRAMEWORK FOR THE PISA CONTEXT QUESTIONNAIRES

For PISA 2012, the conceptual framework for the context questionnaires was published together with the assessment frameworks for mathematics, reading, science, problem solving and financial literacy. Therefore this section provides a summary of the context questionnaire framework only, with the interested reader referred to further details in OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. The framework for the context questionnaires in PISA 2012 outlines how PISA can be developed further as a sustainable database for educational policy and research. To this end, the framework starts with a review of the general purpose and policy relevance of PISA. Three types of policy-relevant "products" are identified:

- ***Indicators*** monitor the functioning, productivity and equity of education systems. PISA-based indicators refer to cognitive outcomes as well as non-cognitive outcomes such as attitudes, beliefs, motivation and learning-related behaviour, the latter being measured within the Student Questionnaire.
- PISA provides ***knowledge*** on individual, school and system-level factors that determine educational effectiveness. The programme reports representative, reliable data on factors that, according to previous research, are expected to impact student achievement. In addition to describing these factors, PISA estimates their direct and indirect relationships to student performance and other outcomes. Thus, it helps to understand how educational outcomes are produced.
- Each PISA assessment updates the sustainable, comparative ***database*** that allows researchers world-wide to study policy-oriented questions. PISA provides a data source for the study of educational contexts in general (e.g. how family, school and out-of-school education interact) and the study of educational variables in economic and sociological contexts (e.g. the relationship between demographics, economic wealth, economic growth and human resources).

Some of the relevant factors in understanding student performance, attitudes, and behaviours, and the functioning of education systems are straightforward (such as demographic variables, previous educational career choices, instructional time, and class size), some have been well established in previous PISA assessments (such as student socio-economic status, cognitive strategies, school-level decision-making), while others have proven to be less easily addressed within the PISA design (e.g. accountability policies at the system level, teacher variables, aspects of the classroom learning environment, or out-of-school activities). Choosing among the many variables that might be incorporated into the design is a complex process, directed by the priorities that countries have set for the study, but also informed by educational research.

In its Chapter 6 section on "The general knowledge base: Research in educational effectiveness", the framework outline shows that the student questionnaire, the school questionnaire and the international options are rooted in well-established research instruments (OECD, 2013). Effectiveness factors can roughly be classified as being either input or

processes. Input factors are mostly related to the individual's social and personal background. Also, structural features like school size and funding are treated as inputs. Processes include learning and teaching as core processes with variables designed to capture their quantity and quality. Moreover, professional activities by teachers and principals as well as school policies and practices are classified as process variables. Figure 3.1 provides an overview of input, process, and outcome factors that are covered in the PISA 2012 Questionnaire Design.

■ Figure 3.1 ■

Taxonomy of educational outcomes and predictive factors

	Input	Processes	Outcomes
Students	Gender, grade level, socio-economic status	Attendance/truancy	Mathematical performance
	Educational career, grades	Outside-class activities - e.g. participation in after school programmes	Mathematics-related attitudes, beliefs and motivation
	Immigration background Family environment and support	Motivation, engagement	General school-related attitudes (towards learning outcomes and activities) and behaviour, e.g. commitment, truancy
	ICT[1] experience, attitudes, skills	Learning and thinking strategies, test taking strategies	Learning motivation
	Openness, perseverance, problem solving styles	Learning time (including homework and private tuition)	
Classrooms	Class size, socio-economic background and ethnic composition	Opportunity to learn: Experience with various kinds of mathematical tasks, Concept familiarity Teaching practices: Teacher-directed instruction, student orientation, formative assessment and feedback Teaching quality: Classroom management/ disciplinary climate, teacher support, cognitive activation	
	Teacher education/training, expertise	Instructional time, grouping practices	
Schools	Socio-economic background and ethnic composition Affluence of the community	Achievement orientation, shared norms, leadership, teacher morale and co-operation, professional development	Promotion/retention and graduation rates
	School funding, public vs. private School size	Admission and recruitment policies, tracking, course offerings/school curriculum, evaluation	Attendance
	Parental involvement	Teacher-student relations	
Countries (Systems)	Economic wealth, social (in)equality	School funding, tracking and allocation, policies for professional teacher development, support for special needs and language minority students, hiring and certification policies	Average graduation level
	Diversity policies	Accountability and evaluation policies, locus of decision-making	

1. Information and Communication Technologies.

As PISA 2012 again has mathematics as its major domain, specific consideration has been given to issues of teaching and learning mathematics. This focus is present in three areas of the questionnaire design, as outlined in the Chapter 6 section of the Assessment Framework titled "*Learning conditions for mathematical literacy*" (OECD, 2013), namely non-cognitive outcomes, explanation of students' intentions and behaviours related to mathematics and classroom teaching.

Non-cognitive outcomes: Measures of intrinsic and instrumental motivation for Mathematics, Learning Strategies (Control vs. Elaboration vs. Memorisation), self-efficacy, self-concept, and mathematics anxiety have been taken up from PISA 2003 after careful re-evaluation of their psychometric qualities.

Explaining student intentions and behaviour related to mathematics: How confident students are about their ability to solve mathematical tasks, as well as how students value mathematics, are highly relevant factors in predicting or explaining student behaviour with regard to mathematics, e.g. course-taking and career decisions. A number of expectancy value models both in psychology and in economics have been proposed to integrate both aspects of decision-making. One such model is Ajzen's (1991) theory of planned behaviour, which states that volitional behaviour is determined by specific attitudes and subjective norms (= value component) plus perceived behavioural control (= expectancy component). In PISA 2012, a version of this model has been implemented in the Student Questionnaire. Students' attitudes and attributions, perceptions of control, and subjective norms may predict their work ethics and intentions – e.g. their desire to spend time on mathematics homework – their study behaviour and finally their mathematics performance.

Classroom Teaching: PISA 2012 aims to identify country (and probably school) level profiles in opportunities to learn. Students were confronted with carefully crafted mathematics tasks – some representing mathematical abilities and content categories as mentioned in the PISA mathematics framework, some representing more traditional tasks asking for procedural and declarative knowledge. Following each of those items, students are asked to judge whether and how often they have seen similar tasks in their mathematics lessons and in previous assessments. These measures of content exposure are complemented by several scales describing teaching practices and teaching quality.

The framework's centrepiece, however, is its aim to map out a design for the PISA context questionnaires that will be sustainable well into the future (see Chapter 6 section "*Specifying the Questionnaire Design for PISA 2012*" in OECD, [2013]). To this end, the framework puts a system in place that accommodates recurring general material that is covered in every cycle and domain-specific material (for mathematics, science, or reading literacy, respectively), that is covered every third cycle, thus allowing for trend analyses of general as well as domain-specific issues. In addition, the framework's system also allows for thematic extensions and specific foci to enable PISA to anticipate and incorporate new material or topics of interest to its audience. The following types of measures are differentiated:

(I) General variables (for all cycles)

- Student-level inputs (grade, gender, parental education and occupation, family wealth, educational resources, cultural possessions, immigration status, heritage language, age on arrival in country, family support).
- School-level contexts and inputs (community size, resources, qualifications of teaching staff).
- School-level processes (decision-making, admission policies, assessment and evaluation policies, professional development, teacher engagement/morale, teacher-student relations, parental involvement).
- Instructional processes (learning time, disciplinary climate, teacher support).
- General non-cognitive outcomes – Commitment to learning (behavioural: truancy; personal goal: educational aspirations; motivational: learning engagement, affective: sense of belonging).

(II) Domain-specific trend variables (for major domain only, included every 9 years)

- Domain-specific non-cognitive outcome variables (strategies and metacognition, domain-related beliefs, self-related beliefs, motivation).
- Domain-specific processes variables (Opportunity To Learn, teaching practices, teaching quality, system- and school-level support).

(III) Thematic extension variables (extensions within individual cycles)

- International options (e.g. in 2012, educational career; ICT familiarity).
- Context variables for additional domains (e.g. ICT-related experiences relevant for computer-based problem solving).
- Descriptive and explanatory variables for specific reports (e.g. in 2012: mathematics-related motivations and intentions based on the theory of planned behavior).
- Malleable variables at the school level (e.g. tracking policies, teacher certification) that are specifically selected for descriptive purposes or for causal inference.

(IV) System-level data, mainly gathered outside of PISA

- Output of educational institutions (e.g. certificates).
- Financial and human resources invested into education.
- Access to and participation in education.
- Learning environment and organisation of schools.

An appropriate balance between (I), (II), (III), and (IV) is considered crucial for the overarching design of PISA questionnaires, and for the long term success of the PISA programme. In order to establish valid and reliable trends at the country level, it is important to implement a constant set of general variables in all cycles both for the calculation of proficiency estimates and as major reporting variables. Thus, these context and input background variables should not change. In order to provide trend information on non-cognitive outcomes and mathematics-related context/process variables, PISA 2012 retained as many variables that were used in the Student and School Questionnaires in 2003 as possible, unless they were shown not to work cross-culturally or not to account for differences in outcomes. Figure 3.2 provides an overview of the mathematics-specific indices in the student questionnaire that provided trend information between 2003 and 2012.

■ Figure 3.2 ■

Student Questionnaire – Mathematics-specific trend scales 2003-2012

Scale name 2003/2012	Scale label
General and mathematics processes	
BELONG*	Sense of belonging to school
STUREL	Student-teacher relations at school
DISCLIM	Disciplinary climate in the mathematics classroom
TEACHSUP	Teacher support in the mathematics classroom
Non-cognitive outcomes – Self and mathematics related cognitions	
ANXMAT	Mathematics anxiety
ATSCHL	Attitudes towards school: Learning outcomes
INSTMOT	Instrumental motivation to learn mathematics
INTMAT	Interest in and enjoyment of mathematics
MATHEFF	Mathematics self-efficacy
SCMAT	Mathematics self-concept

*This scale has been extended from 6 to 9 items in 2012. Trend analyses should only involve the 6 common items (i.e. ST87Q01 to ST87Q06).

CROSS-CULTURAL COMPARABILITY OF MEASURES IN THE CONTEXT QUESTIONNAIRES

One of the major challenges of an international study such as PISA is the cross-cultural validity and applicability of all instruments. In PISA 2012, the phenomenon that a number of non-cognitive student context constructs had been shown to be linked to performance in unexpected ways was given much thought and attention during the development phase of the context questionnaires (Kyllonen, Lietz and Roberts, 2010). More specifically, at the between-country level, data from previous cycles were such that countries with higher performance levels in a subject showed less positive attitudes towards that subject whereas more positive attitudes were recorded for lower-performing countries (Van de gaer and Adams, 2010; Van de gaer et al., 2012). Cross-cultural difference in response styles were considered to be – at least part of – the reason for this phenomenon.

Cross-cultural differences in response styles have been considered to represent a serious source of bias in international surveys that use Likert items. Several types of response styles – including extreme, central, acquiescent and disagreement response styles – have been described (e.g. Greenleaf, 1992; Clarke, 2000; Johnson et al., 2005). All of them can make it difficult to distinguish authentic cultural differences from “stylistic” biases in respondent behaviour (Van de Vijver and Poortinga, 1997; van Hemert, Poortinga and van de Vijver, 2007).

Proposed explanations of differences in response styles include the assumption of frame-of-reference effects whereby responses to attitude (or other) questions might differ systematically depending on which frame of reference (either across countries or across sub-groups within countries) is applied. These frames-of-reference include so-called “cultural macro values” (e.g. Hofstede, 2001; Schwartz, 2006; Triandis et al., 1988), the “Big Fish Little Pond Effect” (for an analysis using PISA 2000 data see Marsh and Hau, 2003), and social desirability (Holtgraves, 2004).

Three approaches, although intertwined, were identified in PISA 2012 to address this phenomenon. First, the phenomenon could be considered to reflect genuine differences between countries whereby some countries or cultural groups might have more positive attitudes regardless of the fact that the related actual context or outcome of interest is worse than in other countries. Second, it could be regarded as a measurement issue in that the measures or item types employed accentuate differences in response styles between countries and cultural groups. Therefore, it would be desirable to pursue measures that would be less affected by different response styles. Third, it could be considered that this phenomenon could be adjusted for through the application of different methods during the analysis stage (see, for example, Van de gaer and Adams, 2010).

In PISA 2012, the second approach was pursued further and four new item formats were introduced to the PISA 2012 Student Questionnaire, namely anchoring vignettes, signal detection debiasing based on the overclaiming technique, forced choice items, and Situational Judgment Tests (SJTs).

Anchoring Vignettes

The first of the new methods was an alternative scoring of Likert-type items based on so-called anchoring vignettes (King and Wand, 2007; Hopkins and King, 2010). The anchoring vignettes approach has been used for making cross-country comparisons in various fields of research (Kapteyn, Smith and Van Soest, 2007; Salomon, Tandon and Murray, 2004; Kristensen and Johansson, 2008) but PISA 2012 was the first educational large-scale assessment to use the technique.

Two sets of so-called anchoring vignettes (see Figures 3.3 and 3.4) were included in the PISA 2012 Student Questionnaire to allow for alternative scoring of self-report items based on students' defined standards when using the 4-point agreement scale (strongly agree – agree – disagree – strongly disagree).

Each of these vignettes described behaviours of a hypothetical mathematics teacher that were indicative of lower or higher levels of classroom management (Figure 3.3) or teacher support (Figure 3.4), respectively. Each vignette combined several behavioural aspects. Students read the vignettes and were asked to indicate their level of agreement with a statement about the hypothetical teachers described in the vignettes. Differences in these ratings could be attributed to differences in the interpretation of the rating scale and general differences in preferred response behaviours as the underlying levels in the hypothetical teachers were held constant across countries.

When items were scored based on vignettes, numerical values for student responses were not assigned based on the concrete response option chosen (e.g., the value 4 for "strongly agree" and 3 for "agree") but based on the self-report answer relative to the personal standard captured by the respondent's individual rating of the three vignettes that form one set. The extension of the nonparametric scoring procedure (e.g., King and Wand, 2007) is described step by step in Chapter 17 of this report.

Clear interpretation of the vignettes in terms of the relative ordering of low, medium, and high levels of the described characteristics was one requirement for the use of vignettes. Results from analysis of Field Trial and Main Survey data showed that the vignettes capturing classroom management behaviours (see Chapter 17) produced clearer results (e.g. regarding the correct rank order of low, medium, and high vignettes by most respondents) and were better suited as anchors for students' self-report answers than the teacher support vignettes. In other words, a higher proportion of students did not give tied responses and the number of order violations – i.e., respondents' evaluations of the three anchors that violated the theoretically expected "correct" order – was lower for the classroom management vignettes than for the teacher support vignettes. These findings indicated that the former vignettes were worded in a way that made the difference between the high and low vignette larger than the latter vignettes.

■ Figure 3.3 ■

Anchoring vignettes based on classroom management behaviours

Low level	The students' in Mr. <name's> class frequently interrupt his lessons. As a result, he often arrives five minutes late to class.	ST84Q03
Medium level	The students' in Ms. <name's> class frequently interrupt her lessons. She always arrives five minutes early to class.	ST84Q01
High level	The students' in Ms. <name's> class are calm and orderly. She always arrives on time to class.	ST84Q02

Note. For each vignette students were asked to indicate how much they agree with the statement "Mr./Ms. <name> is in control of his/her classroom."

■ Figure 3.4 ■

Anchoring vignettes based on teacher support behaviours

Low level	Ms. <name> sets mathematics homework once a week. She never gets the answers back to students before examinations.	ST82Q03
Medium level	Mr. <name> sets mathematics homework once a week. He always gets the answers back to students before examinations.	ST82Q02
High level	Ms. <name> sets mathematics homework every other day. She always gets the answers back to students before examinations.	ST82Q01

Note. For each vignette students were asked to indicate how much they agree with the statement "Mr./Ms. <name> is concerned about his students' learning."

Topic Familiarity with Signal Detection Correction

The PISA 2012 student questionnaire includes several questions regarding familiarity with certain mathematics topics that were designed to measure students' opportunities to learn and content knowledge. When students are asked how well they know a given concept or whether they have seen a certain task type in their mathematics class, responses might, however, be affected by the same response tendencies that were revealed for other constructs.

One possible way of correcting for such response tendencies is the use of the so-called Overclaiming Technique (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy, and Underwood, 1977). This technique is a method that can be used to estimate both respondents' concept familiarity and their tendency to overstate what they know. It does this by collecting recognition judgments for intermixed concepts that actually exist, and foils, i.e. concepts that do not exist. In the PISA 2012 Student Questionnaire (ST62) this was operationalised by asking students to indicate their familiarity – on a 5-point scale from "never heard of it" to "know it well; understand

the concept" – with 13 actual mathematics concepts (e.g. "polynomial function") and three foils (i.e. "proper number", "subjunctive scaling" and "declarative fraction"). Foils were created by combining a term from grammar (i.e. "proper", as in proper noun; "subjunctive", as in subjunctive mood; "declarative" as in declarative sentence) with a mathematical term (i.e. number; scaling; fraction, respectively).

As discussed in Chapter 17, two indices were computed from students' responses to this question (ST62). One index was a simple mean of students' familiarity scores on the 5-point scale with the thirteen actual concepts (FAMCON). The other index took that mean and subtracted from it the mean familiarity score of the three foil concepts (FAMCONC).

Simple indices that can be derived are the so-called "Hit-Rate" and the "False-Alarm Rate". From these, more complex indices of accuracy and bias could be derived based on Signal Detection Theory (SDT) approaches. Figure 3.5 gives an overview of all indices that were compared for the Field Trial. This figure also includes two additional indices, namely "topic familiarity" and "foil familiarity" that were calculated based on Field Trial data. These two additional indices are simple average scores derived from manifest student responses across all 16 items of the test.

■ Figure 3.5 ■

Overview of most prominent Signal Detection Measures and additional scoring rules for PISA 2012 questionnaire items

	Measure	Description/Formula
1	# hits	Number of real items rated as familiar
2	# misses	Number of real items rated as unfamiliar
3	# false alarms	Number of foils rated as familiar
4	# correct rejections	Number of foils rated as unfamiliar
5	H (hit rate)	Proportion of real items rated as familiar
6	F (false-alarm rate)	Proportion of foils rated as familiar
	P(c)	Percent correct (Hits + Correct Rejections)
7	z(H)	z-standardised hit rate
8	z(F)	z-standardised false-alarm rate
9	d′ ("d prime")	The number of hits relative to the number of false-alarms; d′=z(H)-z(F)
10	C (Bias)	-.5* (z(H)=z(F))
11	Topic Familiarity[ł] (FAMCON)	Mean response for all concepts
12	Foil Familiarity[ł]	Mean response for all foils
13	Adjusted Topic Familiarity[ł] (FAMCONC)	Difference score: Topic Familiarity – Foil Familiarity

*Dependent on cut-off value

[ł] alternative indices investigated in PISA Field Trial, not based on SDT

Situational Judgment Tests

Situational Judgment Test items (SJTs; Weekley and Ployhart, 2006) present short descriptions of situation with several possible responses which the test-taker must evaluate. There are many variations, but most often SJT items present several response options, and ask respondents to: (a) select the best option (multiple-choice); or, the best and the worst; (b) indicate for each option whether it would be acceptable or not (true-false), or (c) rate each option using a Likert scale. SJTs are widely used in industry and increasingly in education. In addition to the demonstrated validity of SJTs in employment settings (e.g. see McDaniel et al., 2001), SJTs have been shown as valid predictors in educational contexts such as performance during medical studies as well (e.g. Lievens et al., 2005). SJTs can reduce adverse impacts on, for example, mean score differences between racial groups as they tend to rely less on cognitive abilities than traditional item formats. Therefore, SJTs might be more appropriate instruments for minority groups than traditional tests.

Situational Judgment Tests were applied in the PISA 2012 Field Trial to measure two different constructs, namely Mathematics Motivation, and Problem Solving. Based on Field Trial results, only the Problem Solving SJT was retained for the Main Survey. The Motivation SJT did function reasonably well but could not add validity in terms of increasing hypothesised relationships with other relevant constructs beyond the traditional Likert scales.

The Problem Solving SJT in the PISA 2012 student questionnaire consisted of three different scenarios that described situations that could arise in the course of solving a problem. Questions focus on a person's initial response to a problem as well as possible approaches to take if one's initial response to the problem fails. The three scenarios involved a) a problem with a text message on a mobile phone, b) route selection for getting to a zoo and c) a malfunctioning ticket vending machine. Response options to each scenario tapped into different problem-solving strategies, namely systematic strategies, unsystematic strategies and seeking help.

Forced Choice

If respondents have to decide between different alternatives, many of the problems associated with Likert scales can be avoided. In a so-called Forced-Choice assessment, a respondent is asked to choose one out of several alternative descriptions or assign ranks to all descriptions according to the extent to which the items describe the respondent's personality. Any ranking of *n* items can be coded equivalently using $[n(n\text{-}1)]/2$ binary outcome variables. For $n = 4$ choices {A, B, C, D}, the respondent has to assign ranking positions to each alternative, usually numbers from 1 (most preferred) to 4 (least preferred). The number of data-points that can be generated from this ranking is maximal but the cognitive load of such a comparison is also high. Alternatively, the respondent might be asked to indicate his or her most and least preferred option. This represents a partial ranking because it only assigns the first and the last ranks. The number of data points that can be generated is only somewhat smaller than for the full ranking, at the benefit of a reduction in cognitive load. A third alternative is to ask the test-taker to only choose his or her most (or least) preferred option. The simplest form of Forced Choice item is a paired comparison between only two choices. The ambiguity of the instruction is considerably reduced. Questions such as "Which of the two attributes describes you better?" or "Please rank the following 4 attributes according to how well they describe you" define much more clearly what the test taker has to do than the question "To what extent do you agree with each the following statements?" A drawback is that the cognitive load of the task increases when several response options have to be compared against each other.

Because it is impossible to endorse every item, the forced-choice format eliminates uniform biases such as acquiescence responding (Cheung and Chan, 2002), and can increase operational validity by reducing "halo" effects (Bartram, 2007). Forced Choice methods can reduce (but not fully eliminate) response biases. The reduction of bias is maximal when items in each block do not differ regarding their social desirability and other response styles, such as acquiescence or central tendency (e.g. Brown and Maydeu-Olivares, 2013).

Forced Choice (FC) assessments are more fake-resistant than Likert-type questionnaires. Especially when statements of equal social desirability are compared, faking becomes very complicated. Studies have shown reduced score inflation and maintenance of criterion related validities of FC measures in situations where examinees are motivated to fake (Bowen, Martin, and Hunt, 2002; Christiansen et al., 2005; White and Young, 1998).

Three important recent developments in psychometric models for FC data are the approaches by (a) Steve Stark, (b) Jimmy de la Torre, and (c) Anna Brown. Stark et al. (2005) proposed a model, the *Multi-Unidimensional Pairwise-Preference Model* (MUPP), for constructing and scoring multidimensional pairwise preference items. De la Torre et al. (2011) extended Stark et al.'s (2005) model by suggesting an item response model for preference data that can accommodate more than two components, and also different formats. They illustrate the application of the Bayesian Ipsative Data Analysis (BIDA) algorithm based on the MUPP model using Monte Carlo (MC) simulations. An alternative solution to the problem of ipsative data was presented by Brown and Maydeu-Olivares (2013) who suggest to transform ranking data into a series of paired comparisons. This transformed data can then be analysed with bi-factor models that account for the local independencies between the ranking based pairwise comparisons. The item characteristic function (ICF) for the binary outcome variable y, which is the result of comparing item *i* measuring trait a and item *k* measuring trait *b*, is then simply a standard two-dimensional normal ogive IRT (Item Response Theory) model for binary data with two exceptions: First, factor loadings are structured so that every binary outcome *yl* involving the same item will share the same factor loading. Second, uniquenesses of latent response variables are structured so that they equal the sum of the 2 items involved. Third, the item characteristic functions are not independent, but patterned covariance matrices need to be specified.

A simpler approach is to derive a score for Forced Choice items based on the number of endorsements of one type of statements, i.e. an ipsative scoring strategy. This strategy is obviously inferior to the above described IRT models, but an alternative for small numbers of items, especially when the interest is not in deriving scores for all constructs but just the preference for one specific behaviour or attitude. This principle was used in the PISA 2012 Field Trial to measure students' preferences for Mathematics versus other subjects, as well as for their preferences for certain learning strategies.

TRANSITIONING THE CONTEXT QUESTIONNAIRES FROM PAPER ADMINISTRATION TO ONLINE ADMINISTRATION MODE

In addition to paper-based delivery of the questionnaires, PISA 2012 introduced an online administration mode for the School Questionnaire. On this first occasion, online administration of the School Questionnaire was optional for countries. Within countries, the possibility to print a pdf version of the School Questionnaire was provided, mainly to

enable school principals to obtain information that they had to get from elsewhere to answer some questions (e.g. about staff qualifications).

Nineteen countries and economies took up the online School Questionnaire option in the Main Survey in PISA 2012 which resulted in the administration of the questionnaire in 24 language versions. Participants included: Australia, Austria, Chile, Cyprus,[1] Denmark, Estonia, Finland, Hungary, Iceland, Ireland, Israel, Korea, Liechtenstein, Norway, Portugal, Singapore, Slovenia, Switzerland and Chinese Taipei.

Improvements to the online School Questionnaire from the Field Trial to the Main Survey targeted international contractor processes and questionnaire functionalities for both National Centres and respondents, namely school principals or their deputies.

While the processes for the production of the online School Questionnaire were largely parallel to the processes for the production of paper-based questionnaires, a number of areas required additional work to support the transition to an online mode of administration:

a) The creation of online source versions in English and French of the survey and survey architecture (e.g. online question construction; variable naming; validation rules; administration error messages). Online source versions were produced to:

   - help National Centres in the authoring of national versions of questionnaires; and
   - be incorporated in verification processes along with the negotiated Questionnaire Adaptation Spreadsheet (QAS) against the nationally adapted version of the questionnaire.

b) The development of functionalities for the administration of the online School Questionnaire, which included:

   - import and management of country sampling frameworks in the online management interface; and
   - management of differentiated survey access between Consortium partners, National Centres and participating sampled schools.

c) The development of new validity checks before and after implementation of the Main Survey. Online school questionnaire data were directly exported into *KeyQuest* – the data capture and cleaning software specifically developed for PISA – between consortium partners rather than data being exported from NCs into *KeyQuest* as with paper-based questionnaires These checks included:

   - variable naming checks during the Final Check and linguistic verification processes before Main Survey implementation; and
   - validity reports that were run in *KeyQuest* of school sampling IDs from the online survey management interface.

Additional improvements for the Main Survey administration included: international contractor management of nationally adapted Field Trial questionnaires and Main Survey source updates in the online platform, improvements to the online authoring tool for National Centres, Consortium help and feedback during the authoring process.

UPDATE OF THE INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS (ISCO) FROM ITS 1988 VERSION TO THE 2008 VERSION

Prior to PISA 2012 the 1988 version of the International Standard Classification of Occupations, ISCO-88, was used to code responses to open-ended questions by students about their mother's and father's occupation. In 2007, a new version, namely ISCO-08, was adopted by the International Labor Organisation (ILO) and recommended to be used by both the ILO and the European Commission (2009) in official statistics. The updated version covered more appropriately current occupations, particularly in the area of Information and Communication Technology and also defined more clearly different managerial levels. Hence, it was decided to adopt the ISCO-08 classification in PISA 2012.

In addition to including the nominal four-digit ISCO codes, the PISA dataset also include a mapping of ISCO onto an assumed interval scale – *International Socio-Economic* Index of Occupational Status (ISEI) which has been developed as a scale that is reflective of socio-educational status and is comparable across countries (Ganzeboom, 2010; Ganzeboom and Treiman, 2003). Together with information on parental education and home possessions, ISEI is subsequently used to create the *PISA index of economic, social and cultural status* (ESCS). The rationale for using these three components is that socio-economic status is usually seen to be based on education, occupational status and income.

ESCS is used in many PISA reports and analyses, both as a control for the socio-economic status of students and schools and in bivariate correlations with performance as one of the main indicators of equity in an education system. Hence, the Consortium undertook analyses to examine the impact of the change from ISCO-88 to ISCO-08.

ESCS-88 is used as a label for the ESCS index that involves using ISEI values computed based on the ISCO-88 and ISEI-88. ESCS-08 is used as a label for the ESCS index that involves using ISEI values computed based on the ISCO-08 and ISEI-08.

To support the change from ESCS-88 to ESCS-08 a range of analyses were undertaken to document the implications of the update in terms of means, distributions of ESCS as well as the relationship between ESCS and student performance using Main Survey data from PISA 2012. Secondly, analyses aimed at exploring whether the changes in the ISCO classification have had implications for particular codes using data from the double coding process of the PISA 2012 Field Trial were undertaken. These results are reported in Chapter 17.

THE MEASUREMENT OF OPPORTUNITY TO LEARN

Current research on effective teaching (e.g., the comprehensive review of international teaching effectiveness research written by Good, Wiley and Florez for the 2009 edition of the *International Handbook of Research on Teachers and Teaching*) uses three kinds of measures to describe the classroom learning environment, namely measures of content, teaching practices and teaching quality.

Aspects of content matter, how it is selected, structured, and presented, have often been treated under the heading of Opportunity to Learn (OTL). The breadth and depth of content are described, coherence is rated and the alignment between intended curriculum (i.e. stated standards, syllabi) and implemented curriculum (i.e. the content actually taught) is evaluated. Schmidt and Maier (2009) argue that OTL is a rather straightforward concept: “What students learn in school is related to what is taught”, and they suggest to focus on OTL “in the narrowest sense: Student’s content exposure”.

Another set of measures refers to specific practices that are used by teachers, such as teacher-directed and student-directed activities, or various kinds of assessments. A well-known overview of evidence on teaching practices is provided by Hattie (2008). The OECD’s Teaching and Learning International Survey (TALIS) asked teachers about the frequency of using 13 different teaching practices which could be grouped into three dimensions: structuring practices (e.g. “I explicitly state learning goals.”), student-oriented practices (e.g. “Students work in small groups to come up with a joint solution to a problem or task.”), and enhanced activities (“Students work on projects that require at least one week to complete.”).These three dimensions could be identified across cultures (OECD, 2009).

Third, classroom environments have been characterised by aspects of the quality of teaching, i.e. how teachers deliver content and practices in the classroom. According to Pianta and Hamre (2009) who developed one of the most influential protocols for classroom observations, three dimensions are underlying the quality of teaching, namely classroom organisation, emotional support, and instructional support. This model has gained support from studies of teachers (Tschannen-Moran and Woolfolk-Hoy, 2007) as well as classroom research elsewhere (see Klieme, Pauli and Reusser, 2009; Baumert et al., 2010 who use the term “cognitive activation” rather than “instructional support”).

Opportunity to Learn, teaching practices, and quality measures may be combined to describe and evaluate classroom teaching and learning across cultures. All three kinds of measures have been implemented in the PISA 2012 Student Context Questionnaire to obtain information regarding the learning environment for mathematics.

Sometimes, the label “Opportunity to Learn” is used as embracing all aspects of instruction experienced by the student (e.g., Stevens, 1993). The PISA 2012 Questionnaire Framework, however, defines OTL as “coverage of content categories and problem types” to differentiate it from teaching practices and quality of teaching (OECD, 2013).

In PISA, the measurement of OTL has to be modified from approaches used in other studies, as the mathematics assessment is not framed according to content elements, but refers to fundamental mathematical abilities and broad content categories. Therefore, the measurement of OTL is based mainly on student judgements.

Opportunity to Learn content

Opportunity to Learn – in the sense of mathematical content that students experience – was assessed in PISA 2012 in three ways as detailed below.

Experience with mathematical tasks (ST61)

This question asked students how often they encounter various types of mathematical tasks during their time in school. Here, two subscales were formed from the same list of nine tasks that was used to measure of student self-efficacy in mathematics, a) experience with pure mathematical tasks (EXPUREM: ST61Q05, ST61Q07, ST61Q09), and b) experience with applied mathematical tasks (EXAPPLM: ST61Q01, ST61Q02, ST61Q03, ST61Q04, ST61Q06, ST61Q08).

Familiarity with mathematical concepts (ST62)

This question asked students to judge how familiar they were with 13 mathematical concepts. The response scale had five options: never heard of it (1), heard of it once or twice (2), a few times (3) or often (4), know it well and understand the concept (5).

All 13 items were combined into an overall index, Familiarity with Mathematics Concepts (FAMCON). Based on the 13 items, two indexes were used in reporting (see OECD 2014), a) Index of familiarity with algebra (ST62Q01, ST62Q03, and ST62Q05) and b) Index of Familiarity with Geometry (ST62Q06, ST62Q11, ST62Q13, and ST62Q15) but not included in the international database.

Question ST62 also included three foils, i.e. non-existing pseudo-concepts (ST62Q04, ST62Q10, and ST62Q12). If students indicated they heard of these or even know them well, this indicated overclaiming. The familiarity measure could be adjusted for the tendency to overclaim (see earlier explanation under "Topic Familiarity with Signal Detection Correction"); the adjusted index, Familiarity with Mathematics Concepts – Corrected for Overclaiming (FAMCONC) has been included in the international database.

Exposure to types of mathematical tasks in lessons and in tests (ST73-ST76)

Students were exposed to carefully crafted mathematics tasks – some representing applied mathematical reasoning as assessed in the PISA mathematics test, some representing inner-mathematical reasoning such as proofs and geometrical constructions, some representing short, well-defined word problems as frequently used in textbooks, or tasks checking procedural knowledge. For each of these four types of mathematical tasks, a short characterisation and two examples from different areas of mathematics are provided. Students were instructed not to solve these tasks. Instead, they were asked to recall how often they had previously encountered similar tasks in a) their mathematics lessons and b) in assessments on choosing one of four response options, namely "frequently", "sometimes", "rarely" or "never".

As a result, two variables, one indicating the frequency in mathematics lesson and one indicating the frequency experienced in tests for the following four types of mathematical tasks:

- OTL - Algebraic Word Problem (ST73)
- OTL - Procedural Task (ST74)
- OTL - Pure Math Reasoning (ST75)
- OTL - Applied Math Reasoning (ST76)

Opportunity to Learn teaching practices

To operationalise this component of OTL, the teaching practices items from the OECD TALIS survey were adapted for use in PISA 2012. The items were reframed for use with students and some practices that are specific to mathematics were added. After some items were removed based on results in the Field Trial, 13 teaching practices remained in question ST79 which formed the following three scales:

- Teacher behaviour, Teacher-directed instruction (TCHBEHTD) based on items ST79Q01, ST79Q02, ST79Q06, ST79Q08, ST79Q15;
- Teacher behaviour, Student orientation (TCHBEHSO) based on items ST79Q03, ST79Q04, ST79Q07, ST79Q10; and
- Teacher behaviour, Formative assessment (TCHBEHFA) based on items ST79Q05, ST79Q11, ST79Q12, ST79Q17.

Opportunity to Learn teaching quality

As mentioned above, current research on teaching suggests that (a) classroom organisation and management, (b) teacher emotional and social support, and (c) cognitive activation have to be addressed as basic dimensions of instructional quality.

Two of these dimensions were covered in PISA 2003 and the respective scales continued to be used in 2012:

- Disciplinary climate (DISCLIMA), based on all five items in ST81, indicating *problems* with classroom organisation; and
- Mathematics teaching (TEACHSUP), based on all five items in ST77.

Response options for the items in both scales were “every lesson”, “most lessons”, “some lessons” and “never or hardly ever”.

The third dimension, cognitive activation (COGACT), based on nine items in ST80, is new to PISA 2012. Students were asked the extent to which they felt challenged by the tasks set by their mathematics teacher (e.g., “We usually have to think for a while in order to solve the problems we are assigned by our mathematics teacher”). This scale was used previously as a national option in PISA 2003 in Germany (see Baumert et al., 2008).

To test the usefulness of anchoring vignettes for adjusting non-cognitive scales for cross-cultural differences in response style in survey such as PISA, two scales measuring the quality of mathematics teaching were used.

The first scale, namely Teacher Support (MTSUP), consisted of one new item (ST83Q01 “My teacher lets us know we need to work hard”) plus three of the five items in ST77 that were used in the scale Mathematics teaching (TEACHSUP) but with changed response options, namely “strongly agree”, “agree”, “disagree” and “strongly disagree”. The corresponding anchoring vignette consisted of three items in ST82 that described three teachers in terms of the frequency of setting and returning homework.

The second scale, namely Classroom Management (CLSMAN), consisted of three items that were akin to the items in the Disciplinary climate (DISCLIMA) scale plus one item (ST85Q04) that was taken verbatim from that scale (i.e. ST81Q03). The corresponding anchoring vignette consisted of three items in ST84 that described three teachers of different levels of punctuality for lessons and student behaviour in class. Further details regarding the use of anchoring vignettes have been provided in an earlier section of this chapter.

THE ROTATION OF THE STUDENT CONTEXT QUESTIONNAIRE – DESIGN AND INTENDED ANALYSES

Whereas rotation of cognitive skills tests has been used extensively to increase content coverage of assessed domains for a long time, rotated student context questionnaires were used for the first time in a main data collection of an international comparative assessment in education in PISA 2012. This was done to increase the content coverage of topics of interest to PISA in the questionnaire without increasing the response time for individual students to more than 30 minutes.

The rotated design was such that three forms of the questionnaire contained a common part and a rotated part. The common part, which was administered to all students, contained questions to obtain information about gender, language at home, migrant background, home possessions, parental occupation and education. The rotated part which was administered to one-third of students contained questions about attitudinal and other non-cognitive constructs.

Prior to going down the path of using rotated student questionnaires in the main data collection, extensive analyses were undertaken to examine the impact of this methodology on the continuity of the results. Thus, PISA 2006 data for nine heterogeneous countries were rescaled after having been restructured to simulate the outcomes of the use of different rotated context questionnaire designs. Results revealed negligible differences when means, standard deviations, percentiles were estimated using plausible values drawn with multilevel item response models that adopted different approaches to questionnaire rotation. Also, only 110 of 2 700 correlations between student context constructs and proficiency differed by more than 0.03 with standard errors increasing either not at all or by 0.01 (Adams, Lietz and Berezner, 2013).

The logistics of questionnaire administration became slightly more complex by using a rotated Student Questionnaire design for several reasons. First, more adjustments needed to be negotiated between National Project Managers (NPMs) and the Consortium. Second, although the absolute number of student questionnaires to be printed remained the same for a given sample size, different forms had to be printed, increasing production costs. Third, during administration, about the same number of students had to respond to randomly assigned Student Questionnaire forms which remained

relatively simple as Student Questionnaire forms were not linked to specific cognitive test forms. Despite these slightly more complex logistics, the rotated Student Questionnaire was administered successfully in the great majority of participating countries. This was not least due to the experience with the administration of rotated forms of the Student Questionnaire in all PISA Field Trials to date.

The finally chosen design as illustrated in Figure 3.6 was a rotation with constructs being asked in two of the three forms to allow joint analyses of these constructs. This resulted in responses from two third of students per construct but freed up less space. Still, it was considered preferable as a full covariance matrix could be derived as every construct was asked with every other construct at least once.

■ Figure 3.6 ■

Final design of rotated Student Context Questionnaires in PISA 2012

Form A	Form B	Form C
Common part (8 minutes)		
Rotated question set 1 (11 **minutes**)	Rotated question set 3 (11 **minutes**)	**Rotated question set 3 missing**
Rotated question set 2 (11 **minutes**)	**Rotated question set 3 missing**	Rotated question set 2 (11 **minutes**)
Rotated question set 3 missing	Rotated question set 1 (11 **minutes**)	Rotated question set 3 (11 **minutes**)

Notes: Three rotated forms, two-thirds of students answer questions in rotated parts.

As can be seen in Figure 3.6, the rotated student context questionnaire in PISA 2012 consisted of two parts, namely the "common" and the "rotated" part. Questions in the "common" part were answered by all students while questions in the "rotated" parts were answered by two thirds of the student sample.

It should be noted that each rotated question set occurred first in one of the forms in order to balance the possibility of missing data due to respondents' fatigue in the latter part of the questionnaire. Thus, as can be seen in Figure 3.6, Form A contained question set 1 first, question set 3 features first in Form B while in Form C students were first asked to respond to question set 2.

The common part was estimated to take students about eight minutes to complete. Each rotated question set could take up about eleven minutes of response time. The common part and two rotated question sets, then, resulted in the usual 30 minute response time to the Student Questionnaire for an individual student. Timing estimates were derived from knowledge gained from previous PISA cycles as well as cognitive laboratories during the item development and Field Trial phase.

The content in the common part (see Figure 3.7) included demographics questions and major reporting variables.

For the rotated parts of the Student Questionnaire, the following guiding principles were applied in the allocation of questions to the three question sets:

- Use intact scales only. Do not split items constituting a construct across forms.
- Allocate questions with similar themes to a question set.
- Each question set not to exceed 11 minutes; question sets should be of similar length.
- Balance constructs in terms of their correlation with performance. In other words, on average, correlation with performance of constructs in question sets should be similar based on results of Field Trial.

Three question sets were designed in this way whereby question set 1 was included in Forms A and B. Question set 2 was included in Forms A and C. Question set 3 was included in Forms B and C (see Figure 3.7). Details regarding the questions in the rotated part of the three Student Questionnaire forms are given in Figure 3.9.

Question set 1 contained items covering attitudes towards mathematics and the problem solving Situational Judgement Test items. Question set 2 included items on school climate and attitudes towards school. Mathematics anxiety was also included in question set 2 although, conceptually, it would have been place more appropriately in question set 1. However, as items in question set 1 already showed reasonable correlations with performance while correlations between items and performance were a bit weaker in question set 2, mathematics anxiety was placed in question set 2 due to its relatively higher correlation with mathematics performance. Question set 3 consisted of items measuring Opportunity to Learn and learning strategies. Overall, question 3 was slightly shorter than question sets 1 and 2 but only marginally so.

■ Figure 3.7 ■

Questions in the common part

Question number	Description
ST01	Grade
ST02	Country study programme
ST03	Age of student
ST04	Sex of student
ST05	Attend <ISCED 0>
ST06	Age at <ISCED 1>
ST07	Grade Repeating
ST08	Truancy; Times late for school
ST09	Truancy; Days unexcused absence
ST115	Truancy; Times skipped classes
ST11	Family structure
ST12	Mother's occupation (ISCO); Component of ESCS
ST13	Mother's educational level – Schooling (ISCED); Component of ESCS
ST14	Mother's educational level – Post school (ISCED); Component of ESCS
ST15	Mother's current job status; Component of ESCS
ST16	Father's occupation (ISCO); Component of ESCS
ST17	Father's educational level (ISCED) – Schooling (ISCED); Component of ESCS
ST18	Father's educational level (ISCED) – Post school (ISCED); Component of ESCS
ST19	Father's current job status; Component of ESCS
ST20	Immigrant background
ST21	Age of arrival in test country
ST25	Language spoken at home
ST26	General home possessions plus country-specific wealth items; Component of ESCS
ST27	Number of certain possessions in household; Component of ESCS
ST28	Books at home

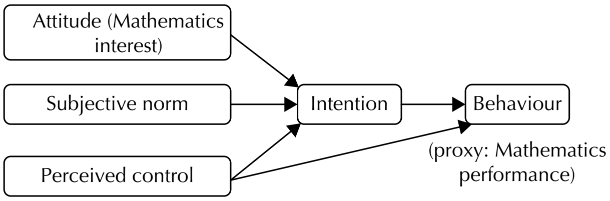
Notes: Questions are listed in the order in which they appeared in the Student Questionnaire. Question numbers in the Field Trial and Main Survey were the same which meant that some numbers are missing (e.g. ST22 to ST24) because they were deleted after the trial from the Student Questionnaire (some countries included these questions in the Educational Career Questionnaire). Also, some numbers were new (e.g. ST115) as some questions were revised substantially after the Field Trial.

Analysts interested in exploring approaches to problem solving in the Student Questionnaire – maybe in order to relate these approaches to proficiency in problem solving from the cognitive tests – are pointed to question set 1 which covers this area (ST94, ST96, ST101 and ST104).

Some scales in the PISA 2012 context questionnaire framework and subsequent questionnaire were designed to enable the exploration of the Theory of Planned Behaviour (Ajzen, 1991; Armitage and Conner, 2001; Sheeran, 2002). Analysts interested in exploring a model derived from this theory such as the one presented below, should also turn to question set 1.

■ Figure 3.8 ■

Model of Theory of Planned Behaviour



■ Figure 3.9 ■

Questions in the rotated parts

FORM A		FORM B		FORM C	
Q. number	**Description**	**Q. number**	**Description**	**Q. number**	**Description**
ST01-28	Common part (see Figure 3.7)	ST01-28	Common part (see Figure 3.7)	ST01-28	Common part (see Figure 3.7)
ST29	Instrumental Motivation (Q2, 5, 7, 8) Mathematics Interest (Q1, 3, 4, 6)	ST42	Mathematics Self-Concept (Q2, 4, 6, 7, 9); Mathematics Anxiety (Q1, 3, 5, 8, 10)	ST53	Learning Strategies (Self-Control)
ST35	Subjective Norms	ST77	Teacher Support in Mathematics Class	ST55	Out-of-School- Lessons
ST37	Mathematics Self-Efficacy	ST79	Teaching Practices	ST57	School Study Time
ST43	Perceived Control of Mathematics Performance	ST80	Cognitive Activation in Mathematics Lessons	ST61	Experience with Applied Maths Tasks (Q1-4, 6, 8); Experience with Pure Math tasks (Q5, 7, 9)
ST44	Attributions to Failure in Mathematics	ST81	Disciplinary Climate	ST62	Familiarity with Maths Concepts
ST46	Mathematics Work Ethic	ST82	Anchoring Vignettes - Teacher Support	ST69	Minutes in <Class Period>
ST48	Mathematics Intentions (Forced-Choice)	ST83	Mathematics Teacher Support	ST70	Number of <Class Period> per Week
ST49	Mathematics Behaviour	ST84	Anchoring Vignettes - Classroom Management	ST71	Number of All <Class Period> per Week
ST93	Perseverance	ST85	Mathematics Teacher's Classroom Management	ST72	Class Size
ST94	Openness for Problem Solving	ST86	Student-Teacher Relations	ST73	OTL - Algebraic Word Problem
ST96	Problem Solving Strategies (SJT-Text Message)	ST87	Sense of Belonging to School	ST74	OTL - Procedural Task
ST101	Problem Solving Strategies (SJT-Route Selection)	ST88	Attitude towards School: Learning Outcomes	ST75	OTL - Pure Mathematics Reasoning
ST104	Problem Solving Strategies (SJT-Ticket Machine)	ST89	Attitude towards School: Learning Activities	ST76	OTL - Applied Mathematics Reasoning
ST53	Learning Strategies (Self-Control)	ST91	Perceived Control of Success in School	ST42	Mathematics Self-Concept (Q2,4,6,7,9); Mathematics Anxiety (Q1,3,5,8,10)
ST55	Out-of-School-Lessons	ST29	Instrumental Motivation (Q2, 5, 7, 8) Mathematics Interest (Q1, 3, 4, 6)	ST77	Teacher Support in Mathematics Class
ST57	School Study Time	ST35	Subjective Norms	ST79	Teaching Practices
ST61	Experience with Applied Mathematics Tasks (Q1-4, 6, 8); Experience with Pure Mathematics tasks (Q5, 7, 9)	ST37	Mathematics Self-Efficacy	ST80	Cognitive Activation in Mathematics Lessons
ST62	Familiarity with Mathematics Concepts	ST43	Perceived Control of Mathematics Performance	ST81	Disciplinary Climate
ST69	Minutes in <Class Period>	ST44	Attributions to Failure in Mathematics	ST82	Anchoring Vignettes - Teacher Support
ST70	Number of <Class Period> per Week	ST46	Mathematics Work Ethic	ST83	Mathematics Teacher Support
ST71	Number of All <Class Period> per Week	ST48	Mathematics Intentions (Forced-Choice)	ST84	Anchoring Vignettes - Classroom Management
ST72	Class Size	ST49	Mathematics Behaviour	ST85	Mathematics Teacher's Classroom Management
ST73	OTL - Algebraic Word Problem	ST93	Perseverance	ST86	Student-Teacher Relations
ST74	OTL - Procedural Task	ST94	Openness for Problem Solving	ST87	Sense of Belonging to School
ST75	OTL - Pure Mathematics Reasoning	ST96	Problem Solving Strategies (SJT-Text Message)	ST88	Attitude towards School: Learning Outcomes
ST76	OTL - Applied Mathematics Reasoning	ST101	Problem Solving Strategies (SJT-Route Selection)	ST89	Attitude towards School: Learning Activities
		ST104	Problem Solving Strategies (SJT-Ticket Machine)	ST91	Perceived Control of Success in School

Note

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

References

Adams, R.J., P. Lietz and **A. Berezner** (2013), "On the use of rotated context questionnaires in conjunction with multilevel item response models", *Large-scale Assessments in Education*, No. 1 (1).

Ajzen, I. (1991), "The theory of planned behaviour", *Organizational Behavior and Human Decision Processes*, No. 50, pp. 179-211.

Armitage, C.J. and **M. Conner** (2001), "Efficacy of the theory of planned behaviour: A meta-analytic review", *British Journal of Social Psychology*, No. 40, pp. 471-499.

Bartram, D. (2007), "Increasing validity with forced-choice criterion measurement formats", *International Journal of Selection and Assessment*, No. 15, pp. 263–272. *http://dx/doi.org/10.1111/j.1468-2389.2007.00386.x*

Baumert, J., M. Kunter, W. Blum, M. Brunner, T.Voss, A.Jordan, U. Klusmann, S. Krauss, M. Neubrand and **Y-M. Tsai** (2010), "Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress", American *Educational Research Journal*, Vol. 47, pp. 133-180.

Baumert, J., W. Blum, M. Brunner, T. Dubberke, A. Jordan, U. Klusmann, S. Krauss, M. Kunter , K. Löwen , N. Neubrand and **Y-M. Tsai** (2008), "Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV)" - Dokumentation der Erhebungsinstrumente (Baumert et al. [2009]), Materialien aus der Bildungsforschung, No. 83, Max-Planck-Institut für Bildungsforschung, Berlin.

Bowen, C.C., B.A. Martin, and **S.T. Hunt** (2002), "A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires*", International Journal of Organizational Analysis*, No. 10(3), pp. 240-259.

Brown, A., and **A. Maydeu-Olivares** (2013), "How IRT can solve problems of ipsative data in forced-choice questionnaires", *Psychological Methods*, No. 18(1), p. 36.

Cheung, M.W.L. and **W. Chan** (2002), "Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis", *Structural Equation Modeling*, No. 9, pp. 55–77.
http://dx/doi.org/10.1207/S15328007SEM0901_4

Christiansen, N., G. Burns and **G. Montgomery** (2005), "Reconsidering the use of forced-choice formats for applicant personality assessment", *Human Performance*, No. 18, pp. 267-307.
http://dx/doi.org/10.1207/s15327043hup1803_4

Clarke, I., III. (2000), "Extreme response style in cross-cultural research: An empirical investigation", *Journal of Social Behavior and Personality*, No. 15, pp. 137-152.

De la Torre, J., V. Ponsoda, I. Leenen and **P. Hontangas** (2011), "Some extensions of the Multi-Unidimensional Pairwise Preference Model", Paper presented at the 26th Annual Conference for the Society of Industrial and Organizational Psychology, Chicago, Illinois.

Ganzeboom, H.B.G. (2010), "Do's and Don't's of Occupation Coding with an Extension on ISCO-08", Working Paper of the Department of Social Research Methodology, VU Amsterdam, last revised January 3, 2010.

Ganzeboom, H. B.G. and **D. J. Treiman** (2003), "Three Internationally Standardised Measures for Comparative Research on Occupational Status", in Hoffmeyer-Zlotnik, J. H.P. and C.Wolf (eds.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Kluwer Academic Press New York, pp. 159-193.

Good, T. L., C. R. H. Wiley and **I. R. Florez** (2009), "Effective teaching: an emerging synthesis", in G. Dworkin (ed.), *International Handbook of Research on Teachers and Teaching*, Springer, New York, pp. 803–816.

Greenleaf, E.A. (1992), "Improving rating scale measures by detecting and correcting bias components in some response styles", *Journal of Marketing Research*, No. 24 , pp. 176-188.

Hattie, J. (2008), *Visible learning: A synthesis of over 800 Meta-analyses Relating to Achievement*, Routledge.

Hofstede, G. (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, Sage, Thousand Oaks, California.

Holtgraves, T. (2004), "Social desirability and self-reports: Testing models of socially desirable responding", *Personality and Social Psychology Bulletin*, No. 30, pp. 161-172.

Hopkins, D.J. and **G. King** (2010), "Improving Anchoring Vignettes: Designing Surveys to Correct for Interpersonal Incomparability", *Public Opinion Quarterly*, Vol. 74, pp. 201-22.

International Labour Organisation (2008), *International Standard Classification of Occupations: ISCO-08*, International Labour Organisation, Geneva.

Johnson, T., P. Kulesa, Y. I. Cho and **S. Shavitt** (2005), "The relation between culture and response styles: Evidence from 19 countries", *Journal of Cross-Cultural Psychology*, No. 36, 2, pp. 264-277.

Kapteyn, A., J.P. Smith, A.Van Soest and **H. Vonkova** (2007), "Vignettes and Self-Reports of Work Disability in the US and the Netherlands", *American Economic Review*, Vol. 97, pp. 461–73.

King, G. and **J. Wand** (2007), "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes", *Political Analysis*, Vol. 15, No. 1, pp. 46–66.

Klieme, E., C.Pauli and **K. Reusser** (2009), "The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Classrooms", in Janik T. and T. Seidel (eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*, Waxmann Verlag, Münster, pp. 137-160.

Kristensen, N. and **E. Johansson** (2008), "New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes", *Labour Economics*, Vol. 15, pp. 96-117.

Kyllonen, P., Lietz, P., and **Roberts, R.** (2010), "Cross-cultural comparative questionnaire issues", Paper prepared for the PISA2012 Questionnaire Expert Group (QEG) meeting, 25-27 February 2010, Hong Kong.

Lievens, F., T. Buyse and **P.R. Sackett** (2005), "The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains", *Journal of Applied Psychology*, No. 90, pp. 442–452

Marsh, H.W. and **K.T. Hau** (2003), "Big-Fish-Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools", *American Psychologist*, Vol. 58(5), May 2003, pp. 364-376.
http://dx/doi.org/10.1037/0003-066X.58.5.364

McDaniel, M.A., F.P. Morgeson, E.B. Finnegan, M.A. Campion and **E.P. Braverman** (2001), "Predicting job performance using situational judgment tests: A clarification of the literature", *Journal of Applied Psychology*, No. 86, pp. 730–740.

OECD (2014), *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris, pp. 167-200.
http://dx.doi.org/10.1787/9789264190511-en

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS,* OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264068780-en

Pianta, R.C. and **B.K. Hamre** (2009), "Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity", *Educational Researcher,* No. 38, pp. 109–119.
http://dx/doi.org/10.3102/0013189X09332374

Paulhus, D. L., P. D.Harms , N. M. Bruce and **D.C. Lysy,** (2003), "The over-claiming technique: Measuring self-enhancement independent of ability*", Journal of Personality and Social Psychology*, Vol. 84(4), April 2003, pp. 890-904.
http://dx/doi.org/10.1037/0022-3514.84.4.890

Salomon, J. A., A. Tandon and **C.J.L. Murray** (2004), "Comparability of Self-Rated Health: Cross-Sectional Multi-Country Survey Using Anchoring Vignettes", *British Medical Journal*, Vol. 328, pp. 258–61.

Schwartz, S.H. (2006), "Value orientations: Measurement, antecedents and consequences across nations", in Jowell, R., C. Roberts, R. Fitzgerald and G. Eva (eds.), *Measuring Attitudes Cross-nationally - Lessons from the European Social Survey*, Sage, London.

Sheeran, P. (2002), "Intention-behaviour relations: a conceptual and empirical review", in Strobe W. and M. Hewstone (eds.), *European Review of Social Psychology*, Vol. 12, Wiley, Chichester, pp. 1-30.

Schmidt, W.H. and **A. Maier** (2009), "Opportunity to Learn", in Sykes G., B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, Routledge, New York, pp. 541-559.

Stark, S., O.S. Chernyshenko and **F. Drasgow,** (2005), "An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model", *Applied Psychological Measurement*, No. 29 (3), pp. 184-203.

Stevens, F. (1993), "Applying an Opportunity-to-Learn Conceptual Framework to the Investigation of the Effects of Teaching Practices via Secondary Analyses of Multiple-Case-Study Summary Data", *Journal of Negro Education*, Vol. 62, (3), pp. 232-248.

Tschannen-Moran, M. and **A. Woolfolk-Hoy** (2007), "The differential antecedents of self-efficacy beliefs of novice and experienced teachers", *Teaching and Teacher Education*, Vol. 23, pp. 944-956.

Triandis, H.C., R.Bontempo, M.J.Villareal, M. Asai and **N. Lucca** (1988), "Individualism and Collectivism: Cross-Cultural Perspectives on Self - Ingroup Relationships", *Journal of Personality and Social Psychology*, No. 54, (2), pp. 323-338.

Van de gaer, E. and **R.J. Adams** (2010), "*The Modeling of Response Style Bias: An Answer to the Attitude-Achievement Paradox?"*, Paper presented at the annual meeting of the American Education Research Association (AERA), 30 April - 4 May, 2010, Denver, Colorado.

Van de gaer, E., A.Grisay, W.Schulz and **E. Gebhardt** (2012), "The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries", *Journal of Cross-Cultural Psychology* , No. 43(8), pp. 1205-1228

Van de Vijver, F.J.R. and **Y.H. Poortinga,** (1997), "Towards an integrated analysis of bias in cross-cultural assessment", *European Journal of Psychological Assessment*, Vol. 13, (1), Hogrefe and Huber Publishers.

Van Hemert, D., Y.H. Poortinga and **F.J.R. Van de Vijver** (2007), "Emotion and culture: A meta-analysis", *Cognition and Emotion*, No. 21, pp. 913-943.

Weekley J.A. and **R.E. Ployhart** (2006), "*Situational Judgment tests: Theory, Management, and Application"*, Erlbaum, Mahwah, New Jersey.

White, L.A. and **M.C. Young,** (1998), "Development and validation of the Assessment of Individual Motivation (AIM)", Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, California.

Zimmerman, J., P.K.Broder, J.J.Shaughnessy and **B.J. Underwood** (1977), "A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition", *Intelligence*, No. 1, pp. 5-31.

4

Sample Design

TARGET POPULATION AND OVERVIEW OF THE SAMPLING DESIGN

The international PISA target population in each participating country and economy consisted of 15-year-old students attending educational institutions in grade 7 and higher. This meant that countries were to include 15-year-old students:

- enrolled full-time in educational institutions;
- enrolled in educational institutions who attended only on a part-time basis;
- enrolled in vocational training programmes, or any other related type of educational programmes; and
- attending foreign schools within the country (as well as students from other countries attending any of the programmes in the first three categories).

It was recognised that no testing of 15-year-olds schooled full-time in the home, workplace or out of the country would occur and therefore these 15-year-olds were not included in the international target population.

The operational definition of an age population directly depends on the testing dates. The international requirement was that the assessment had to be conducted during a 42-day period, referred to as the testing period, between 1 March 2012 and 31 August 2012, unless otherwise agreed.

Further, testing was not permitted during the first six weeks of the school year because of a concern that student performance levels may have been lower at the beginning of the academic year than at the end of the previous academic year, even after controlling for age.

The 15-year-old international target population was slightly adapted to better fit the age structure of most of the Northern Hemisphere countries. As the majority of the testing was planned to occur in April, the international target population was consequently defined as all students aged from 15 years and 3 completed months to 16 years and 2 completed months at the beginning of the assessment period. This meant that in all countries testing in April 2012, the target population could have been defined as all students born in 1996 who were attending an educational institution as defined above.

A variation of up to one month in this age definition was permitted. This allowed a country testing in March or in May to still define the national target population as all students born in 1996. If the testing was to take place at another time until the end of August, the birth date definition had to be adjusted so that in all countries the target population was always students aged 15 years and 3 completed months to 16 years and 2 completed months at the time of testing, or a one month variation of this.

In all but one country, the Russian Federation, the sampling design used for the PISA assessment was a two-stage stratified sample design. The first-stage sampling units consisted of individual schools having 15-year-old students. Schools were sampled systematically from a comprehensive national list of all PISA-eligible schools, known as the school sampling frame, with probabilities that were proportional to a measure of size. The measure of size was a function of the estimated number of PISA-eligible 15-year-old students enrolled in the school. This is referred to as systematic Probability Proportional to Size (PPS) sampling. Prior to sampling, schools in the sampling frame were assigned to mutually exclusive groups based on school characteristics called explicit strata, formed in particular to improve the precision of sample-based estimates.

The second-stage sampling units in countries using the two-stage design were students within sampled schools. Once schools were selected to be in the sample, a complete list of each sampled school's 15-year-old students was prepared. For each country a Target Cluster Size *(TCS)* was set, this value was typically 35 students although with agreement countries could use alternative values. From each list of students that contained more than the *TCS*, a sample of typically 35 students were selected with equal probability and for lists of fewer than the *TCS*, all students on the list were selected.

For countries participating in the international option of Financial Literacy (FL), the *TCS* was increased in each sampled school so as to also achieve the required student sample size for FL.

In the Russian Federation, a three-stage design was used. In this case, geographical areas were sampled first (first-stage units) using PPS sampling, and then schools (second-stage units) were selected within these sampled geographical areas. Students were the third-stage sampling units in this three-stage design and were sampled from the selected schools.

Population coverage, and school and student participation rate standards

To provide valid estimates of student achievement, the sample of students had to be selected using established and professionally recognised principles of scientific sampling, in a way that ensured representation of the full target population of 15-year-old students in the participating countries.

Furthermore, quality standards had to be maintained with respect to: *(i)* the coverage of the PISA international target population; *(ii)* accuracy and precision; and *(iii)* the school and student response rates.

Coverage of the PISA international target population

National Project Managers (NPMs) might have found it necessary to reduce their coverage of the target population by excluding, for instance, a small, remote geographical region due to inaccessibility, or a language group, possibly due to political, organisational or operational reasons, or special education needs students. In an international survey in education, the types of exclusion must be defined consistently for all participating countries and the exclusion rates have to be limited. Indeed, if a significant proportion of students were excluded, this would mean that survey results would not be deemed representative of the entire national school system. Thus, efforts were made to ensure that exclusions, if they were necessary, were minimised according to the PISA 2012 Technical Standards (see Annex F).

Exclusion can take place at the school level (exclusion of entire schools) or at the within-school level (exclusion of individual students). Areas deemed to be part of a country (for the purpose of PISA), but which were not included for sampling, although this occurred infrequently, were designated as non-covered areas. Care was taken in this regard because, when such situations did occur, the national desired target population differed from the international desired target population.

International within-school exclusion rules for students were specified as follows:

- Intellectually disabled students are students who have a mental or emotional disability and who, in the professional opinion of qualified staff, are cognitively delayed such that they cannot be validly assessed in the PISA testing setting. This category includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students were not to be excluded solely because of poor academic performance or normal discipline problems.
- Functionally disabled students are students who are permanently physically disabled in such a way that they cannot be validly assessed in the PISA testing setting. Functionally disabled students who could provide responses were to be included in the testing.
- Students with insufficient assessment language experience are students who need to meet all of the following criteria: *i)* are not native speakers of the assessment language(s); *ii)* have limited proficiency in the assessment language(s); and *iii)* have received less than one year of instruction in the assessment language(s). Students with insufficient assessment language experience could be excluded.
- Students not assessable for other reasons as agreed upon. A nationally-defined within-school exclusion category was permitted if agreed upon by the PISA Consortium. A specific sub-group of students (for example students with dyslexia, dysgraphia, or dyscalculia) could be identified for whom exclusion was necessary but for whom the previous three within-school exclusion categories did not explicitly apply, so that a more specific within-school exclusion definition was needed.
- Students whose language of instruction for mathematics (the major domain for 2012), was one for which no PISA assessment materials were available. Standard 2.1 of the PISA 2012 Technical Standards (see Annex F) notes that the PISA test is administered to a student in a language of instruction provided by the sampled school to that sampled student in the major domain of the test. Thus, if no test materials were available in the language in which the sampled student is taught, the student was excluded.

A school attended only by students who would be excluded for intellectual, functional or linguistic reasons was considered a school-level exclusion.

It was required that the overall exclusion rate within a country (i.e. school-level and within-school exclusions combined) be kept below 5% of the PISA desired target population. Guidelines for restrictions on the level of exclusions of various types were as follows:

- School-level exclusions for inaccessibility, feasibility, or reasons other than those described in the next points were to cover less than 0.5% of the total number of students in the international target population for participating countries. Schools on the school sampling frame which had only one or two PISA-eligible students were not allowed to be

excluded from the frame. However, if, based on the frame, it was clear that the percentage of students in these small schools would not cause a breach of the 0.5% allowable limit, then such schools could be excluded in the field at that time of the assessment, if they still only had one or two PISA-eligible students.

- School-level exclusions for intellectually or functionally disabled students, or students with insufficient assessment language experience, were to cover fewer than 2% of students.
- Because definitions of within-school exclusions could vary from country to country, NPMs were asked to adapt the international definitions to make them workable in their country but still to code them according to the PISA international coding scheme. Within-school exclusions for intellectually disabled or functionally disabled students, or students with insufficient assessment language experience, or students nationally-defined and agreed upon for exclusion were expected to cover fewer than 2.5% of students. Initially, this could only be an estimate. If the actual percentage was ultimately greater than 2.5%, the percentage was re-calculated without considering students excluded because of insufficient assessment language experience since this is known to be a largely unpredictable part of each country's PISA-eligible population, not under the control of the education system. If the resulting percentage was below 2.5%, the exclusions were regarded as acceptable.

Accuracy and precision

A minimum of 150 schools had to be selected in each country; if a participating country had fewer than 150 schools then all schools were selected. Within each participating school, a predetermined number of students, denoted as *TCS* (usually 35 students), were randomly selected with equal probability, or in schools with fewer than *TCS* eligible students, all students were selected. In total, a minimum sample size of 4 500 assessed students was to be achieved, or the full population if it was less than this size. It was possible to negotiate a *TCS* that differed from 35 students, but if it was reduced then the sample size of schools was increased beyond 150, so as to ensure that at least 4 500 students would be assessed. The *TCS* selected per school had to be at least 20 students, so as to ensure adequate accuracy in estimating variance components within and between schools – a major analytical objective of PISA.

NPMs were strongly encouraged to identify available variables to use for defining the explicit and implicit strata for schools to reduce the sampling variance. See later section on stratification for other benefits.

For countries which had participated in previous PISA assessments that had larger than anticipated sampling variances associated with their estimates, recommendations were made about sample design changes that would possibly help to reduce the sampling variances for PISA 2012. These included modifications to stratification variables, and increases in the required sample size.

School response rates

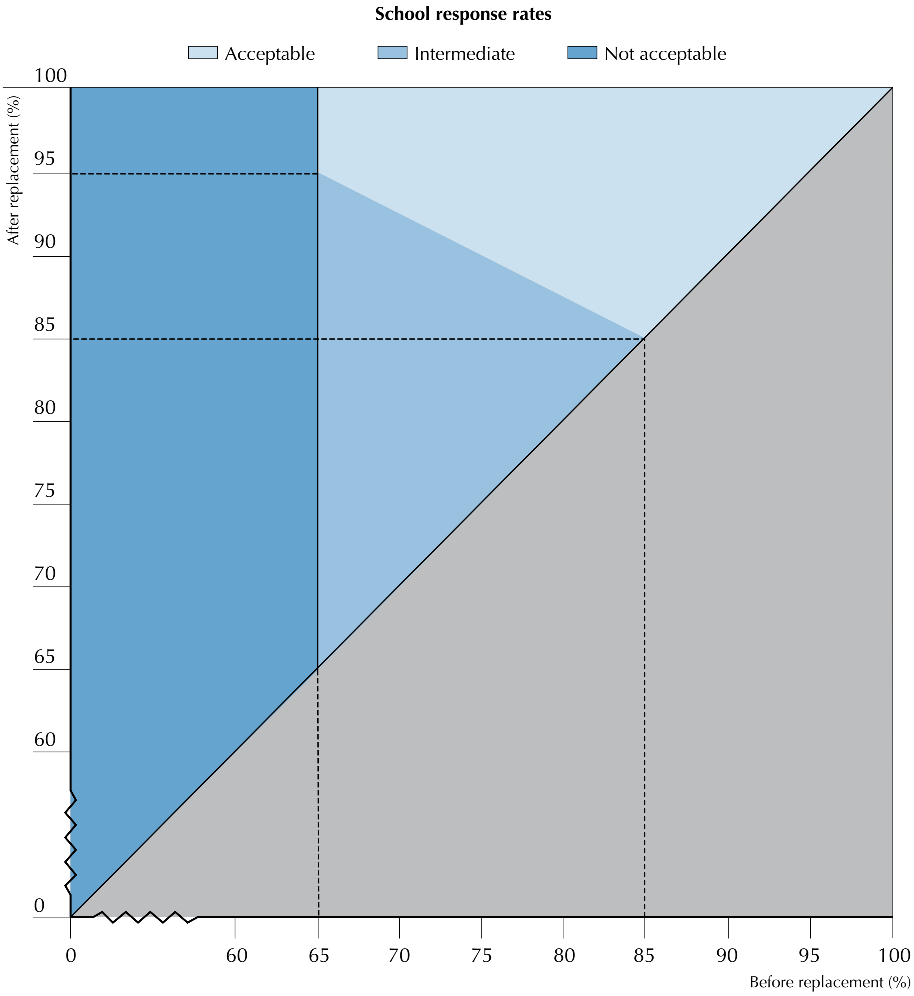
A response rate of 85% was required for initially selected schools. If the initial school response rate fell between 65% and 85%, an acceptable school response rate could still be achieved through the use of replacement schools. Figure 4.1 provides a summary of the international requirements for school response rates. To compensate for a sampled school that did not participate, where possible, two potential replacement schools were identified. Furthermore, a school with a student participation rate between 25% and 50% was not considered as a participating school for the purposes of calculating and documenting response rates.[1] However, data from such schools were included in the database and contributed to the estimates included in the initial PISA international report. Data from schools with a student participation rate of less than 25% were not included in the database, and such schools were regarded as non-respondents.

The rationale for this approach was as follows. There was concern that, in an effort to meet the requirements for school response rates, a National Centre might accept participation from schools that would not make a concerted effort to have students attend the assessment sessions. To avoid this, a standard for student participation was required for each individual school in order that the school be regarded as a participant. This standard was set at a minimum of 50% student participation. However, there were a few schools in many countries that conducted the assessment without meeting that standard. Thus a judgement was needed to decide if the data from students in such schools should be used in the analyses, given that the students had already been assessed. If the students from such schools were retained, non-response bias would possibly be introduced to the extent that the students who were absent could have been different in achievement from those who attended the testing session, and such a bias is magnified by the relative sizes of these two groups. If one chose to delete all assessment data from such schools, then non-response bias would be introduced to the extent that the school was different from others in the sample, and sampling variance would be increased because of sample size attrition.

The judgement was made that, for a school with between 25% and 50% student response, the latter source of bias and variance was likely to introduce more error into the study estimates than the former, but with the converse judgement for those schools with a student response rate below 25%. Clearly the cut-off of 25% is arbitrary as one would need extensive studies to try to establish this cut-off empirically. However, it is clear that, as the student response rate decreases within a school, the possibility of bias from using the assessed students in that school will increase, while the loss in sample size from dropping all of the students in the school will be small.

■ Figure 4.1 ■

School response rate standards



These PISA standards applied to weighted school response rates. The procedures for calculating weighted response rates are presented in Chapter 11. Weighted response rates weigh each school by the number of students in the population that are represented by the students sampled from within that school. The weight consists primarily of the enrolment size of 15-year-old students in the school, divided by the selection probability of the school. Because the school samples were selected with PPS, in most countries many schools contributed equal weights, and as a consequence the weighted and unweighted school response rates were similar. Exceptions could occur in countries that had explicit strata that were sampled at very different rates.

Student response rates

An overall weighted response rate of 80% of selected students in participating schools was required. A student who had participated in the original or follow-up cognitive sessions was considered to be a participant. A minimum student response rate of 50% within each school was required for a school to be regarded as participating: the overall student

response rate was computed using only students from schools with at least a 50% student response rate. Again, weighted student response rates were used for assessing this standard. Each student was weighted by the reciprocal of his/her sample selection probability.

MAIN SURVEY SCHOOL SAMPLE

Definition of the national target population

NPMs were first required to confirm their dates of testing and age definition with the PISA Consortium. NPMs were warned to avoid having any possible drift in the assessment period lead to an unapproved definition of the national target population.

Every NPM was required to define and describe their country's target population and explain how and why it might deviate from the international target population. Any hardships in accomplishing complete coverage were specified, discussed and approved or not, in advance. Where the national target population deviated from full coverage of all PISA-eligible students, the deviations were described and enrolment data provided to measure the degree to which coverage was reduced. The population, after all exclusions, corresponded to the population of students recorded on each country's school sampling frame. Exclusions were often proposed for practical reasons such as increased survey costs or complexity in the sample design and/or difficult test conditions. These difficulties were mainly addressed by modifying the sample design to reduce the number of such schools selected rather than to exclude them. Schools with students that would all be excluded through the within-school exclusion categories could be excluded up to a maximum of 2% as previously noted. Otherwise, countries were instructed to include the schools but to administer the PISA UH (une heure) booklet (see Chapter 2 for more details on the UH booklet), consisting of a subset of the PISA assessment items, deemed more suitable for students with special education needs. Eleven countries used the UH booklet for PISA 2012.

Within participating schools, all PISA-eligible students (i.e., born within the defined time period and in Grade 7 or higher) were to be listed. From this, either a sample of students equal in size to the *TCS* was randomly selected or all students were selected if there were fewer students than the *TCS*. The lists had to include students deemed to meet any of the categories for exclusion, and a variable maintained to briefly describe the reason for exclusion. This made it possible to estimate the size of the within-school exclusions from the sample data.

It was understood that the exact extent of within-school exclusions would not be known until the within-school sampling data were returned from participating schools, and sampling weights computed. Participating country projections for within-school exclusions provided before school sampling were known to be estimates.

NPMs were made aware of the distinction between within-school exclusions and nonresponse. Students who could not take the PISA achievement tests because of a permanent condition were to be excluded and those with a temporary impairment at the time of testing, such as a broken arm, were treated as non-respondents along with other sampled students who were absent.

Exclusions by country are documented in Chapter 11.

The sampling frame

All NPMs were required to construct a school sampling frame to correspond to their national defined target population. The school sampling frame was defined in the *School Sampling Preparation Manual*[2] as a frame that would provide complete coverage of the national defined target population without being contaminated by incorrect or duplicate entries or entries referring to elements that were not part of the defined target population. It was expected that the school sampling frame would include any school that could have 15-year-old students, even those schools which might later be excluded, or deemed ineligible because they had no PISA-eligible students at the time of data collection. The quality of the sampling frame directly affects the survey results through the schools' probabilities of selection and therefore their weights and the final survey estimates. NPMs were therefore advised to be diligent and thorough in constructing their school sampling frames.

All but one country used school-level sampling frames as their first stage of sample selection. The *School Sampling Preparation Manual* indicated that the quality of sampling frames for both two- and three-stage designs would largely depend on the accuracy of the approximate enrolment of 15-year-olds available (*ENR*) for each first-stage sampling unit.

A suitable *ENR* value was a critical component of the sampling frames since selection probabilities were based on it for both two- and three-stage designs. The best *ENR* for PISA was the number of currently enrolled 15-year-old students. Current enrolment data, however, were rarely available at the time of school sampling, which meant using alternatives. Most countries used the first-listed available option from the following list of alternatives:

- student enrolment in the target age category (15-year-olds) from the most recent year of data available;
- if 15-year-olds tend to be enrolled in two or more grades, and the proportions of students who are aged 15 in each grade are approximately known, the 15-year-old enrolment can be estimated by applying these proportions to the corresponding grade-level enrolments;
- the grade enrolment of the modal grade for 15-year-olds; and
- total student enrolment, divided by the number of grades in the school.

The *School Sampling Preparation Manual* noted that if reasonable estimates of *ENR* did not exist or if the available enrolment data were out of date, schools might have to be selected with equal probabilities which might require an increased school sample size. However, no countries needed to use this option.

Besides *ENR* values, NPMs were instructed that each school entry on the frame should include at minimum:

- school identification information, such as a unique numerical national identification, and contact information such as name, address and phone number; and
- coded information about the school, such as region of country, school type and extent of urbanisation, which could possibly be used as stratification variables.

As noted, a three-stage design and an area-level (geographic) sampling frame could be used where a comprehensive national list of schools was not available and could not be constructed without undue burden, or where the procedures for administering the test required that the schools be selected in geographic clusters. As a consequence, the area-level sampling frame introduced an additional stage of frame creation and sampling (first stage) before actually sampling schools (second stage with the third stage being students). Although generalities about three-stage sampling and using an area-level sampling frame were outlined in the *School Sampling Preparation Manual* (for example that there should be at least 80 first-stage units and at least 40 needed to be sampled), NPMs were also informed that the more detailed procedures outlined there for the general two-stage design could easily be adapted to the three-stage design. The NPM using a three-stage design was also asked to notify the PISA Consortium and received additional support in constructing and using an area-level sampling frame. The only country that used a three-stage design was the Russian Federation, where a national list of schools was not available. The use of the three-stage design allowed for school lists to be obtained only for those areas selected in stage one rather than for the entire country.

Stratification

Prior to sampling, schools were to be ordered, or stratified, in the sampling frame. Stratification consists of classifying schools into *like* groups according to selected variables referred to as stratification variables. Stratification in PISA was used to:

- improve the efficiency of the sample design, thereby making the survey estimates more reliable;
- apply different sample designs, such as disproportionate sample allocations, to specific groups of schools, such as those in particular states, provinces, or other regions;
- ensure all parts of a population were included in the sample; and
- ensure adequate representation of specific groups of the target population in the sample.

There were two types of stratification utilised: explicit and implicit. Explicit stratification consists of grouping schools into strata that will be treated independently from one another or as if they were separate school sampling frames. Examples of explicit stratification variables could be states or regions of a country. Implicit stratification consists essentially of sorting the schools uniquely within each explicit stratum by a set of designated implicit stratification variables. Examples of implicit stratification variables could be type of school, degree of urbanisation, or minority composition. This type of stratification is a way of ensuring a strictly proportional sample allocation of schools across all implicit strata. It can also lead to improved reliability of survey estimates, provided that the implicit stratification variables being considered are correlated with PISA achievement at the school level (Jaeger, 1984). Guidelines were provided in the *Sampling Guidelines FT12 Manual*[3] on choosing stratification variables that would possibly improve the sampling.

■ Figure 4.2 [Part 1/2] ■

Stratification variables used in PISA 2012

		Explicit stratification variables	Number of explicit strata	Implicit stratification variables
OECD	Australia	State/Territory (8); Sector (3); Certainty Selections	25	Geographic Zone (3); School Gender Composition (3); School Socio-economic Level (6); Numeracy Achievement Level (6); ISCED Level (3)
	Austria	Programme (17)	17	School Type (4); Region (9); Percentage of Girls (5)
	Belgium	Region (3); Form of Education – Flemish Community (5), French Community (3), German Community (2); Funding – Flemish Community (2), French Community and German Community (1); ISCED Level – Flemish Community and French Community (3), German Community (1); Educational Tracks – French Community (3), German Community and Flemish Community (1)	29	Grade Repetition – Flemish Community and French Community (5), German Community (1); Percentage of Girls – Flemish Community and French Community (4), German Community (1); School Type – French Community (4), German Community and Flemish Community (1)
	Canada	Province (10); Language (3); School Size (16); Certainty Selections	49	Urbanicity (3); Funding (2); ISCED Level (4)
	Chile	Funding type (3); School level (3); School track (4)	18	Percentage of Girls (6); Urbanicity (2); Region (4)
	Czech Republic	Programmes (6); Region (15); School Size (3)	81	School Size (3); Region for Programmes 3, 4, 5, 6 (15); School Gender Composition (3)
	Denmark	Immigrant Levels (5); Certainty Selections	6	School Type (8); ISCED Level (4); Urbanicity (6); Region (6)
	Estonia	Language (3); Certainty Selections	4	School Type (3); Urbanicity (2); County (15); Funding (2)
	Finland	Region (6); Urbanicity (2); Immigrant Levels (3); Certainty Selections	18	School Type (7)
	France	School Type (4); School Size (4)	6	School Type for small school strata (4); Funding (2)
	Germany	School Category (3); State, for normal schools (16)	18	State for other schools (17); School Type (6)
	Greece	Region (15); Funding (2)	16	School Type (3); Funding (2)
	Hungary	School Type (6)	6	Region (7); Mathematics Performance (6)
	Iceland	Region (9); School Size (4)	32	Urbanicity (2); ISCED Level (2)
	Ireland	School Size (3); School Type (4); Project Maths Pilot School (1); Non-aided school (1)	11	Socio-Economic Status Category (5); School Gender Composition Category (5)
	Israel	Language and Apprenticeship or not (3); School Orientation (3); Subsectors for Arabic (3); Gender (3)	12	ISCED Level (4); Group Size (3); SES (4); District (3)
	Italy	Region (21); Study Programme (5); Certainty Selections	104	Funding (2)
	Japan	Funding (2); School Type (2)	4	Levels of proportion of students taking University/ College Entrance Exams (4)
	Korea	School Level (2); School Type (2)	3	Urbanicity (3); School Gender Composition (3)
	Luxembourg	School Type (6)	6	School Gender Composition (3)
	Mexico	State (32); School Size (3); Certainty Selections	97	School Level (2); School Programme (7); Funding (2); Urbanicity (2)
	Netherlands	School Track (4)	4	Programme Category (7)
	New Zealand	School Size (3); Certainty Selections	4	School Decile (4); Funding (2); School Gender Composition (3); Urbanicity (2)
	Norway	School Level (3)	3	None
	Poland	School Type (4)	4	School Sub-type (2); Funding (2); Locality (4); Gender Composition (3)
	Portugal	Geographic Region (30); Certainty Selections	31	ISCED Level (3); Funding (2); Urbanicity (3)
	Slovak Republic	School Type (3); Region (8)	24	Sub-type (6); Language (3); Grade Repetition Level (25); Exam (11)
	Slovenia	Programme/Level (7)	7	Location/Urbanicity (5); Gender (3)
	Spain	Region (18); Funding (2); Linguistic Model for the Basque region (4); Certainty Selections	41	None
	Sweden	Funding (2); ISCED Level (2); Urbanicity (6)	12	Geographic LAN (22); Responsible Authority (4); Level of Immigrants (5); Income Quartiles (5)
	Switzerland	Language (3); School has Grade 9 or not (2); Canton (26); Public/Private (2); School Type (4); Certainty Selections	30	School Type (28); Canton (26)
	Turkey	Region (12); Programme Type (4)	38	School Type (18); Gender (3); Urbanicity (2); Funding (2)
	United Kingdom (excluding Scotland)	Country (3); School Type (4); Region – England (4), Northern Ireland (5), Wales (3); Certainty Selections	30	School Gender Composition (3); School Performance – England and Wales (6), Northern Ireland (1); Local Authority – England (151), Wales (22), Northern Ireland (1)
	United Kingdom (Scotland)	Funding (2); School Attainment (6)	9	Gender (3); Area Type (6)
	United States	Region (4); Funding (2)	8	Grade Span (5); Urbanicity (4); Minority Status (2); Gender (3); State (51)

■ Figure 4.2 [Part 2/2] ■

Stratification variables used in PISA 2012

		Explicit stratification variables	Number of explicit strata	Implicit stratification variables
Partners	Albania	Region (3); Urbanicity (2); Funding (2); Certainty Selections	13	ISCED2/Mixed (2)
	Argentina	Area (6)	6	Funding (2); Education type (3); Education level (9); Urbanicity (2); Secular/Religious (2)
	Brazil	State (27); Maintenance (3); Certainty Selections	81	Administration (3); DHI Quintiles (6); ISCED level (4); Urbanicity (2)
	Bulgaria	Region (11)	11	Type of School (8); Size of Settlement (5); Funding (3)
	Colombia	Region (6); Certainty Selections	7	Urbanicity (2); Funding (2); Weekend school or not (2); Gender (5); ISCED Programme Orientation (4)
	Costa Rica	School Type (5); Certainty Selections	6	Programme (2); Urbanicity (2); Shift (2); Region (27); ISCED Level (3)
	Croatia	Dominant Programme Type (6); Certainty Selections	7	Gender (3); Urbanicity (3); Region (6)
	Cyprus[1, 2]	ISCED Programme Orientation (3); Funding (2); Urbanicity (2)	8	Language (2); ISCED Level (3)
	Hong Kong-China	Funding (4)	4	Student Academic Intake (4)
	Indonesia	Indonesia (1)	1	Province (32); Funding (2); School Type and Level (5); National Exam Result (3)
	Jordan	School Type / Funding (7); Certainty Selections	8	Urbanicity (2); Gender (3); Level (2); Shift (2)
	Kazakhstan	Region (16); Language (13); Certainty Selections	59	Urbanicity (2); ISCED Level (3); ISCED Programme Orientation (2); Funding (2)
	Latvia	Urbanicity (4); Certainty Selections	5	School Type/Level (5)
	Liechtenstein	Liechtenstein (1)	1	Funding (2)
	Lithuania	Urbanicity (4); School Type (4); Certainty Selections	17	Funding (2)
	Macao-China	School Type (3); Programme (2); Language (5)	10	Gender (3); School Orientation (2); ISCED Level (2)
	Malaysia	School Category (6)	6	School Type (16); Urbanicity (2); State (16); Gender (3); ISCED Level (2)
	Montenegro	Programme (4); Region (3)	11	Gender (3)
	Peru	Funding (2); Urbanicity (2)	4	Region (26); Gender (3); School Type (7)
	Qatar	School Type (6)	6	Gender (3); Language (2); Level (5); Funding (2); Programme Orientation (3)
	Romania	Programme (2)	2	Language (2); Urbanicity (2); LIC Type (3)
	Russian Federation	Region (42)	42	Location/Urbanicity (9); School Type (8); School Sub-type (5);
	Serbia	Primary/Other (2); Region (6); School Type (4); Certainty Selections	17	Region (5); Programme (7)
	Shanghai-China	ISCED Level (4); ISCED Programme Orientation (2); Selectivity (3); Certainty Selections	6	Urbanicity (2); Funding (2); Vocational School Type (4)
	Singapore	Funding (2); School Level (2); Certainty Selections	4	Gender (3)
	Chinese Taipei	School type (7); Funding (2); Location (2); Certainty Selections	29	County/City area (22); School Gender (3)
	Thailand	Administration (7); School Type (3); Certainty Selections	17	Region (9); Urbanicity (2); Gender (3)
	Tunisia	Geographical Area (6); Urbanicity (3)	18	ISCED Level (3); Funding (2); Percentage of Repeaters (3)
	United Arab Emirates	Emirate (7); Curriculum (5); Funding (2); Certainty Selections	9	School Level (3); School Gender (3)
	Uruguay	Institutional Sector (4); School Level (3); Certainty Selections	11	Location/Urbanicity (4); Gender (4)
	Viet Nam	Broad Geographical Region (3); Funding (2); Urbanicity (3)	16	Economic Region (8); Province (63); School Type (6); Study Commitment (2)

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Figure 4.2 provides the explicit stratification variables used by each country, as well as the number of explicit strata found within each country. For example, Australia had eight explicit strata using states/territories which were then further delineated by three sectors and also had one explicit stratum for certainty selections, so that there were 25 explicit strata in total. Variables used for implicit stratification and the respective number of levels can also be found in Figure 4.2.

As the sampling frame was always finally sorted by school size, school size was also an implicit stratification variable, though it is not listed in Figure 4.2. The use of school size as an implicit stratification variable provides a degree of control over the student sample size so as to possibly avoid the sampling of too many relatively large schools or too many relatively small schools. A variable used for stratification purposes is not necessarily included in the PISA data files.

Assigning a measure of size to each school

For the probability proportional to size sampling method used for PISA, a Measure of Size (*MOS*) derived from *ENR* was established for each school on the sampling frame. *MOS* was generally constructed as: *MOS* = max(*ENR, TCS*). This differed slightly in the case of small schools treatment, discussed later.

Thus, the measure of size was equal to the enrolment estimate (*ENR*), unless enrolment was less than the *TCS*, in which case the measure of size was set equal to the target cluster size. In most countries, the *MOS* was equal to *ENR* or 35 students, whichever was larger.

As schools were sampled with probability proportional to size, setting the measure of size of small schools to 35 students was equivalent to drawing a simple random sample of small schools. That is, small schools had an equally likely chance of being selected to participate.

Countries participating in the PISA 2012 Financial Literacy (FL) option required a proportional increase to their initial *TCS* based on the expected booklet allocation rate within schools. It was expected for these countries that on average, 8 in 43 students would receive financial literacy booklets and 35 in 43 students would receive PISA booklets. Thus, an initial *TCS* of 35 was increased to a *TCS* of 43, and the *MOS* was then equal to *ENR* or 43 students, whichever was larger.

School sample selection

School sample allocation over explicit strata

The total number of schools to be sampled in each country needed to be allocated among the explicit strata so that the expected proportion of students in the sample from each explicit stratum was approximately the same as the population proportions of PISA-eligible students in each corresponding explicit stratum. There were two exceptions. If very small schools required under-sampling, students in them had smaller percentages in the sample than in the population. To compensate for the resulting loss of sample, the large schools had slightly higher percentages in the sample than the corresponding population percentages. The other exception occurred if only one school was allocated to any explicit stratum. In this case, two schools were allocated for selection in the stratum to aid with variance estimation.

Sorting the sampling frame

The *School Sampling Preparation Manual* indicated that, prior to selecting schools, schools in each explicit stratum were to be sorted by variables chosen for implicit stratification and finally by the *ENR* value within each implicit stratum. The schools were first to be sorted by the first implicit stratification variable, then by the second implicit stratification variable within the levels of the first implicit stratification variable, and so on, until all implicit stratification variables were used. This gave a cross-classification structure of cells, where each cell represented one implicit stratum on the school sampling frame. The sort order was alternated between implicit strata, from high to low and then low to high, etc., through all implicit strata within an explicit stratum.

Determining which schools to sample

The PPS-systematic sampling method used in PISA first required the computation of a sampling interval for each explicit stratum. This calculation involved the following steps:

- recording the total measure of size, *S,* for all schools in the sampling frame for each specified explicit stratum;
- recording the number of schools, *D*, to be sampled from the specified explicit stratum, which was the number allocated to the explicit stratum;
- calculating the sampling interval, *I*, as follows: $I = S/D$; and
- recording the sampling interval, *I*, to four decimal places.

If any school in the stratum had a measure of size as large as or larger than the stratum sampling interval, that school became a school selected with certainty. Such a school was removed from its original explicit stratum and placed in a certainty stratum. The four steps above were then repeated in the original stratum, with now (*S* - certainty school measure of size) in the first step, and (*D*-1) schools for the second step. This process continues until there are no more schools selected with certainty.

Next, a random number had to be generated for each explicit stratum. The generated random number (*RN*) was from a uniform distribution between zero and one and was to be recorded to four decimal places.

The next step in the PPS selection method in each explicit stratum was to calculate selection numbers – one for each of the schools to be selected in the explicit stratum. Selection numbers were obtained using the following method:

- Obtaining the first selection number by multiplying the sampling interval, *I*, by the random number, *RN*. This first selection number was used to identify the first sampled school in the specified explicit stratum.
- Obtaining the second selection number by adding the sampling interval, *I*, to the first selection number. The second selection number was used to identify the second sampled school.
- Continuing to add the sampling interval, *I*, to the previous selection number to obtain the next selection number. This was done until all specified line numbers (1 through *number of schools to be sampled*) had been assigned a selection number.

Thus, the first selection number in an explicit stratum was $RN \times I$, the second selection number was $(RN \times I) + I$, the third selection number was $(RN \times I) + I + I$, and so on.

Selection numbers were generated independently for each explicit stratum, with a new random number generated for each explicit stratum.

Identifying the sampled schools

The next task was to compile a cumulative measure of size in each explicit stratum of the school sampling frame that assisted in determining which schools were to be sampled. Sampled schools were identified as follows.

Let *Z* denote the first selection number for a particular explicit stratum. It was necessary to find the first school in the sampling frame where the cumulative *MOS* equalled or exceeded *Z*. This was the first sampled school. In other words, if C_s was the cumulative *MOS* of a particular school *S* in the sampling frame and $C_{(s-1)}$ was the cumulative *MOS* of the school immediately preceding it, then the school in question was selected if: C_s was greater than or equal to *Z*, and $C_{(s-1)}$ was strictly less than *Z*. Applying this rule to all selection numbers for a given explicit stratum generated the original sample of schools for that stratum.

Box 4.1 **Illustration of probability proportional to size (PPS) sampling**

To illustrate these steps, suppose that in an explicit stratum in a participant country, the PISA-eligible student population is 105 000, then:

- the total measure of size, *S*, for all schools is 105 000;
- the number of schools, *D*, to be sampled is 150;
- calculating the sampling interval, *I*, 105 000/150 = 700;
- generate a random number, *RN*, 0.3230;
- the first selection number is 700 × 0.3230 = 226. This first selection number is used to identify the first sampled school in the specified explicit stratum;
- the second selection number is 226 + 700 = 926. The second selection number was used to identify the second sampled school; and
- the third selection number is 926 + 700 = 1 626. The third selection number was used to identify the third sampled school, and so on until the end of the school list is reached. This will result in a school sample size of 150 schools.

The table below also provides these example data. The school that contains the generated selection number within its cumulative enrolment is selected for participation.

Table 4.1 Examples of PPS sampling

School	*MOS*	Cumulative *MOS* (C_s)	Selection number	School selection
001	550	550	226	Selected
002	364	914		
003	60	974	926	Selected
004	93	1067		
005	88	1155		
006	200	1355		
007	750	2105	1626	Selected
008	72	2177		
009	107	2284		
010	342	2626	2326	Selected
011	144	2770		
…	…	…	…	…

Identifying replacement schools

Each sampled school in the Main Survey was assigned two replacement schools from the school sampling frame, if possible, identified as follows. For each sampled school, the schools immediately preceding and following it in the explicit stratum, which was ordered within by the implicit stratification, were designated as its replacement schools. The school immediately following the sampled school was designated as the first replacement and labelled R_1, while the school immediately preceding the sampled school was designated as the second replacement and labelled R_2. The *School Sampling Preparation Manual* noted that in small countries, there could be problems when trying to identify two replacement schools for each sampled school. In such cases, a replacement school was allowed to be the potential replacement for two sampled schools (a first replacement for the preceding school, and a second replacement for the following school), but an actual replacement for only one school. Additionally, it may have been difficult to assign replacement schools for some very large sampled schools because the sampled schools appeared close to each other in the sampling frame. There were times when it was only possible to assign a single replacement school when two consecutive schools in the sampling frame were sampled, or none for the second of three consecutive sampled schools. That is, no unsampled schools existed between sampled schools.

Exceptions were allowed if a sampled school happened to be the last school listed in an explicit stratum. In this case the two schools immediately preceding it were designated as replacement schools. Similarly, for the first school listed in an explicit stratum, in which case the two schools immediately following it were designated as replacement schools.

Assigning school identifiers

To keep track of sampled and replacement schools in the PISA database, each was assigned a unique, three-digit school code and two-digit stratum code (corresponding to the explicit strata) sequentially numbered starting with one within each explicit stratum. For example, if 150 schools are sampled from a single explicit stratum, they are assigned identifiers from 001 to 150. First replacement schools in the Main Survey are assigned the school identifier of their corresponding sampled schools, incremented by 300. For example, the first replacement school for sampled school 023 is assigned school identifier 323. Second replacement schools in the Main Survey are assigned the school identifier of their corresponding sampled schools, but incremented by 600. For example, the second replacement school for sampled school 136 took the school identifier 736.

Tracking sampled schools

NPMs were encouraged to make every effort to confirm the participation of as many sampled schools as possible to minimise the potential for non-response biases. They contacted replacement schools after all contacts with sampled schools were made. Each sampled school that did not participate was replaced if possible. If both an original school and a replacement participated, only the data from the original school were included in the weighted data provided that at least 50% of the PISA-eligible, non-excluded students had participated. If this was not the case, it was permissible for the original school to be labelled as a nonrespondent and the replacement school as the respondent, provided that the replacement school had at least 50% of the PISA-eligible, non-excluded students as participants.

Special school sampling situations

Treatment of small schools

In PISA, schools were classified as very small, moderately small or large. A school was classified as large if it had an *ENR* above the *TCS* (35 students in most countries). A moderately small school had an *ENR* in the range of one-half the *TCS* to *TCS* (18 to 35 students in most countries). A very small school had an *ENR* less than one-half the *TCS* (17 students or fewer in most countries). New for PISA 2012, very small schools were further classified as very small schools with an *ENR* of zero, one, or two students and very small schools with an *ENR* greater than two students but less than one-half the *TCS*. Unless they received special treatment in the sampling, the occurrence of small schools in the sample will reduce the sample size of students for the national sample to below the desired target because the within-school sample size would fall short of expectations. A sample with many small schools could also be an administrative burden with many testing sessions with few students. To minimise these problems, procedures were devised for managing small schools in the sampling frame.

To balance the two objectives of selecting an adequate sample of small schools but not too many small schools so as to hurt student yield, a procedure was recommended that assumed the underlying idea of under-sampling the very small schools by a factor of two (those with an *ENR* greater than two but less than one-half the *TCS*) and under-sampling the

very small schools with zero, one, or two students by a factor of four and to proportionally increasing the number of large schools to sample. To determine whether very small schools should be undersampled and if the sample size needed to be increased to compensate for small schools, the following test was applied.

- If the percentage of students in very small schools ($ENR<TCS/2$) was 1% or MORE, then very small schools were undersampled and the school sample size increased.
- If the percentage of students in very small schools ($ENR<TCS/2$) was LESS than 1% and the percentage of students in moderately small schools ($TCS/2<ENR<TCS$) was 4% or MORE, then there was no required undersampling of very small schools but the school sample size was increased.

If none of these conditions were true, then the small schools contained such a small proportion of the PISA population that they were unlikely to reduce the sample below the desired target. In this case, no undersampling of very small schools was needed nor an increase to the school sample size to compensate for small schools.

If the number of very small schools was to be controlled in the sample without creating explicit strata for these small schools, this was accomplished by assigning a measure of size (*MOS*) of $TCS/2$ to those very small schools with an *ENR* greater than two but less than $TCS/2$ and a measure of size equal to the $TCS/4$ for the very small schools with an *ENR* of zero, one, or two. In effect, very small schools with a measure of size equal to $TCS/2$ were under-sampled by a factor of two (school probability of selection reduced by half), and the very small schools with a measure of size equal to $TCS/4$ were under-sampled by a factor of four (school probability of selection reduced by three-fourths). This was accomplished as follows.

The formulae below assume an initial target school sample size of 150 and a target student sample size of 5 250.

- *Step 1:* From the complete sampling frame, find the proportions of total *ENR* that come from very small schools with *ENR* of zero, one or two ($P1$), very small schools with *ENR* greater than two but fewer than $TCS/2$ ($P2$), moderately small schools (Q), and large schools (R). Thus, $P1 + P2 + Q + R = 1$
- *Step 2:* Calculate the value L, where $L = 1.0 + 3(P1)/4 + (P2)/2$. Thus L is a positive number slightly more than 1.0.
- *Step 3:* The minimum sample size for large schools is equal to $150 \times R \times L$, rounded up to the nearest integer. It may need to be enlarged because of national considerations, such as the need to achieve minimum sample sizes for geographic regions or certain school types.
- *Step 4:* Calculate the mean value of *ENR* for moderately small schools (*MENR*), and for very small schools (*V1ENR and V2ENR*). *MENR* is a number in the range of *TCS/2* to *TCS*, *V2ENR* is a number larger than two but no greater than *TCS/2*, and *V1ENR* is a number in the range of zero to two.
- *Step 5:* The number of schools that must be sampled from the moderately small schools is given by: $(5\,250 \times Q \times L)/(MENR)$.
- *Step 6:* The number of schools that must be sampled from the very small schools (type *P2)* is given by: $(2\,625 \times P2 \times L)/(V2ENR)$.
- Step 7: The number of schools that must be sampled from the very small schools (type *P1)* is given by: $(1\,313 \times P1 \times L)/(V1ENR)$.

To illustrate the steps, suppose that in a participant country, the *TCS* is equal to 35 students, with 10% of the total enrolment of 15-year-olds in moderately small schools, and 5% in each type of very small schools, *P1* and *P2*. Suppose that the average enrolment in moderately small schools is 25 students, in very small schools (type *P2*) it is 12 students, and in very small schools (type *P1*) it is 1.5 students.

- *Step 1:* The proportions of total *ENR* from very small schools is $P1=0.05$ *and* $P2 = 0.05$, from moderately small schools is $Q = 0.1$, and from large schools is $R = 0.8$. It can be shown that $0.05 + 0.05 + 0.1 + 0.8 = 1.0$.
- *Step 2:* Calculate the value L: $L= 1.0 + 3(0.05)/4 + (0.05)/2$. Thus $L = 1.0625$.
- *Step 3:* The minimum sample size for large schools is equal to $150 \times 0.8 \times 1.0625 = 127.5$. That is, at least 128 (rounded up to the nearest integer) of the large schools must be sampled.
- *Step 4:* The mean value of *ENR* for moderately small schools (*MENR*) is given in this example as 25, very small schools of type *P2* (*V2ENR*) as 12, and very small schools of type *P1* (*V1ENR*) as 1.5.
- *Step 5:* The number of schools that must be sampled from the moderately small schools is given by $(5\,250 \times 0.1 \times 1.0625)/25 = 22.3$. At least 23 (rounded up to the nearest integer) moderately small schools must be sampled.

- *Step 6:* The number of schools that must be sampled from the very small schools (type *P2*) is given by (2 625 × 0.05 × 1.0625)/12 = 11.6. At least 12 (rounded up to the nearest integer) very small schools of type *P2* must be sampled.
- *Step 7:* The number of schools that must be sampled from the very small schools (type *P1*) is given by (1 313 × 0.05 × 1.0625)/1.5 = 46.5. At least 47 (rounded up to the nearest integer) very small schools of type *P1* must be sampled.

Combining these different sized school samples gives a total sample size of 128 + 23 + 12 + 47 = 210 schools. Before considering school and student non-response, the larger schools will yield an initial sample of approximately 128 × 35 = 4 480 students. The moderately small schools will give an initial sample of approximately 23 × 25 = 575 students, very small schools of type *P2* will give an initial sample size of approximately 12 × 12 = 144 students, and very small schools of type *P1* will give an initial sample size of approximately 47 × 1.5 = 70.5, i.e., 71 students. The total initial sample size of students is therefore 4 480 + 575 + 144 + 71 = 5 270.

This procedure, called small school analysis, was done not just for the entire school sampling frame, but for each individual explicit stratum. An initial allocation of schools to explicit strata provided the starting number of schools and students to project for sampling in each explicit stratum. The small school analysis for a single unique explicit stratum indicated how many very small schools of each type (assuming under-sampling, if needed), moderately small schools and large schools would be sampled in that stratum. Together, these provided the final sample size, *n*, of schools to select in the stratum. Based on the stratum sampling interval and random start, large, moderately small, and very small schools were sampled in the stratum, to a total of *n* sampled schools. Because of the random start, it was possible to have more or less than expected of the very small schools of either type, *P1* or *P2*, of the moderately small schools, and of the large schools. The total number of sampled schools however was fixed at *n*, and the number of expected students to be sampled was always approximate to what had been projected from the unique stratum small school analysis.

Sampling for Problem Solving assessment (PS) component

Forty-four countries and economies participated in the Problem Solving assessment (PS) conducted via computer: Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Colombia, Croatia, Cyprus,[4] the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hong Kong-China, Hungary, Ireland, Israel, Italy, Japan, Korea, Macao-China, Malaysia, Montenegro, the Netherlands, Norway, Poland, Portugal, the Russian Federation, Serbia, Shanghai-China, Singapore, the Slovak Republic, Slovenia, Spain, Sweden, Chinese Taipei, Turkey, the United Arab Emirates, the United Kingdom (England only), the United States, and Uruguay. Of these 44 countries and economies, 32 also did the optional Computer-Based Assessment of Literacies in Mathematics and Reading (CBAL). Collectively, the computer-based assessment for PISA 2012 (Problem Solving-only or Problem Solving plus CBAL) was referred to as "CBA". When a country participated in CBA, it was expected that CBA student sampling would occur in every PISA sampled and participating school.

Students selected for the computer-based assessments were a subsample of those selected for the paper-based assessments. The overall sample size requirement was 2 100 CBA students for countries that did only Problem Solving (PS) and 2 700 CBA students for countries that did both PS and CBAL. The recommended CBA Target Cluster Size (ETCS) was 14 students per sampled school for PS-only and 18 students per sampled school for PS and CBAL. Unlike the computer-based assessments in the previous PISA cycle, a student's valid participation in the PISA 2012 CBA did not require their participation in the main paper-based PISA assessment. At least 70% of the CBA subsample within each school was expected to participate in the CBA in order to achieve reliable CBA achievement score generation. The CBA student subsample was selected at the same time that the PISA student sample was selected in each school by the student sampling software, *KeyQuest*.

The actual CBA student sample size at each school was calculated with *KeyQuest*, as the minimum of the ETCS, and the number of sampled PISA students. Arrangements had to be made at the school level to either bring in laptops, or to have extra sessions to alleviate any computer resource problems.

If a country had a large PISA school sample and wished to subsample the PISA sampled schools where CBA student sampling would be done, this became an additional national option. Three countries, Brazil, Italy and Spain, chose to have schools subsampled for CBA from their large national school sample. The schools for CBA were subsampled from sampled schools in each explicit stratum. The number to subsample for CBA in each stratum was based on how many schools would have been needed from each explicit stratum for a school sample of 150 schools. Any schools selected

with certainty for the large national school sample and placed in their own stratum, were added back to their original strata for the subsampling of CBA schools. Then schools that were certainty schools for the large national school sample were subsampled with Probability Proportional to Size, while schools that were sampled initially with a probability of less than 1.0 were subsampled with equal probability. The respective probabilities were calculated in such a way that CBA sample of schools was selected overall with Probability Proportional to Size.

PISA and national study overlap control

The Main Survey for PISA 2012 and a national (non-PISA) study were to occur at approximately the same time in some participating countries. Because of the potential for increased burden, an overlap control procedure was used for two countries (Austria and the United States) who requested for there to be a minimum incidence of the same schools being sampled for both PISA and their national (non-PISA) study. This overlap control procedure required that the same school identifiers be used on the PISA and the national study school frames for the schools in common across the two assessments.

The national study samples were usually selected before the PISA samples. Thus, for countries requesting overlap control, the national study centre supplied the PISA Consortium with their school frames, national school IDs, each school's probability of selection, and an indicator showing which schools had been sampled for the national study.

Sample selections for PISA and the national study could totally avoid overlap of schools if schools which would have been selected with high probability for either study had their selection probabilities capped at 0.5. Such an action would make each study's sample slightly less than optimal, but this might be deemed acceptable when weighed against the possibility of low response rates due to the burden of participating in two assessments. This was not requested for PISA 2012. Therefore, if any schools had probabilities of selection greater than 0.5 on either study frame, these schools had the possibility to be selected to be in both studies.

There were also two other occurrences of overlap control. In the case of Colombia, after sample selection, a requirement for an oversample for a particular region was newly identified. A new stratum had to be added for that region for new sample selection. The sample had to be reselected for the stratum which had previously had those region's schools, as well as for all other region strata. Maximum overlap control ensured that all previously sampled schools not from that region would be sampled again (data collection had already started), along with the new sample for the newly stratified region. The other occurrence of overlap control involved the Russian Federation PISA sample. The Russian Federation had one adjudicated region for PISA 2012 which was separately sampled. The main Russian Federation sample also had several schools sampled from that region. To avoid having the same schools selected for the region in the two samples, minimal overlap control was applied.

To control overlap of schools between PISA and another sample, the sample selection of schools for PISA adopted a modification of an approach due to Keyfitz (1951), based on Bayes Theorem. To use PISA and ICCS (the International Civics and Citizenship Study, a concurrent international study conducted at the time of PISA 2009) in an example of the overlap control approach to minimise overlap, suppose that *PROBP* is the PISA probability of selection and *PROBI* is the ICCS probability of selection. Then a conditional probability of a school's selection into PISA (*CPROB)* is determined as follows:

4.1

$$CPROB = \begin{cases} \max\left[0, \left(\dfrac{PROBI + PROBP - 1}{PROBI}\right)\right] & \textit{if the school was an ICCS school} \\ \min\left[1, \dfrac{PROBP}{(1 - PROBI)}\right] & \textit{if the school was not an ICCS school} \\ PROBP \quad \textit{if the school was not an ICCS eligible school} & \end{cases}$$

Then a conditional *CMOS* variable was created to coincide with these conditional probabilities as follows:

$$CMOS = CPROB \times \text{stratum sampling interval}$$

The PISA school sample was then selected using the line numbers created as usual (see earlier section), but applied to the cumulated *CMOS* values (as opposed to the cumulated *MOS* values). Note that it was possible that the resulting PISA sample size could be slightly lower or higher than the originally assigned PISA sample size, but this was deemed acceptable.

Monitoring school sampling

For PISA 2012, as in the previous cycles, it was a strong recommendation that the PISA Consortium select the school samples rather than the participating countries. This was again incorporated into the 2012 procedures to alleviate the weighting difficulties caused by receiving school sampling frame files in many different formats. Japan was the only participant that selected their own school sample, doing so for reasons of confidentiality.

Sample selection for Japan was replicated by the PISA Consortium to ensure quality in this case. All other participating countries' school samples were selected by and checked in detail by the PISA Consortium. To enable this, all countries were required to submit sampling information on forms associated with the following various sampling tasks:

- time of testing and age definition for both the Field Trial and Main Survey were captured on Sampling Task 1 at the time of the Field Trial, with updates being possible before the Main Survey;
- information about stratification for the Field Trial and for the Main Survey was recorded on Sampling Task 2;
- forms or data associated with Sampling Tasks 3, 4, 5 and 6 were all for the Field Trial;
- the national desired target population information for the Main Survey was captured on the form associated with Sampling Task 7a;
- information about the defined national target population was recorded on the form associated with Sampling Task 7b;
- the description of the sampling frame was noted on the form associated with Sampling Task 8a; and
- the school sampling frame was created in one spreadsheet and the list of any excluded schools in a second spreadsheet associated with Sampling Task 8b.

The PISA Consortium completed school sampling and, along with the school sample, returned other information (small school analyses, school allocation, and a spreadsheet that countries could use for tracking school participation). Figure 4.3 provides a summary of the information required for each sampling task and the timetables (which depended on national assessment periods).

■ Figure 4.3 ■

Schedule of school sampling activities

Activity	Submit to Consortium	Due Date
Update time of testing and age definition of population to be tested	Sampling Task 1 – time of testing and age definition	Update what was submitted at the time of the Field Trial, two months before the school sample is to be selected
Finalise explicit and implicit stratification variables	Sampling Task 2 – stratification and other information	Update what was submitted at the time of the Field Trial, two months before the school sample is to be selected
Define national desired target population	Sampling Task 7a – national desired target population	Submit two months before the school sample is to be selected
Define national defined target population	Sampling Task 7b – national defined target population	Submit two months before the school sample is to be selected
Create and describe sampling frame	Sampling Task 8a – sampling frame description	Submit two months before the school sample is to be selected
Submit sampling frame	Sampling Task 8b – sampling frame (in one Excel® sheet), and excluded schools (in another Excel® sheet)	Submit two months before the school sample is to be selected
Decide how to treat small schools; finalise sample size requirements	Sampling Task 9 – treatment of small schools; sample allocation by explicit strata	The Consortium will complete and return this information to the NPM about one month before the school sample is to be selected
Describe population within strata	Population counts by strata	The Consortium will complete and return this information to the NPM when the school sample is sent to the NPM
Select the school sample	Sampling Task 10 – school sample selection	The Consortium will return the sampling frame to the NPM with sampled schools and their replacement schools identified and with PISA IDs assigned when the school sample is selected
Review and agree to the sampling form required as input to *KeyQuest*	Sampling Task 11 – reviewing and agreeing to the Sampling Form for *KeyQuest* (SFKQ)	Countries had one month after their sample was selected to agree to their SFKQ
Submit sampling data	Sampling Task 12 – school participation information and data validity checks	Submit within one month of the end of the data collection period

Once received from each participating country, each set of information was reviewed and feedback was provided to the country. Forms were only approved after all criteria were met. Approval of deviations was only given after discussion and agreement by the PISA Consortium. In cases where approval could not be granted, countries were asked to make revisions to their sample design and sampling forms and resubmit.

Checks that were performed in the monitoring of each set of information are described in the following text. All entries were observed in their own right but those below were additional matters explicitly examined.

As part of the initial pre-form checks, all special situations known about the participating country were verified with the country. Such special situations included, *TCS* values different from 35 students, whether or not the computer-based assessment was being conducted and if so, its ETCS value, whether or not the financial literacy assessment was being conducted, whether or not overlap control procedures with a national (non-PISA) survey were required, whether or not there was any regional or other type of oversampling, whether or not the UH (one hour) booklet would be used, and whether or not any grade or other type of student sampling would be used. Additionally, any country with fewer than 4 500 or just over 4 500 assessed students in either PISA 2006 or 2009 had increased school sample sizes discussed and agreed upon. Additionally, countries which had too many PISA 2009 exclusions were warned about not being able to exclude any schools in the field for PISA 2012. Finally, any country with effective student sample sizes less than 400 in PISA 2009 also had increased school sample sizes discussed and agreed upon.

Sampling Task 1: Time of testing and age definition

- Assessment dates had to be appropriate for the selected target population dates.
- Assessment dates could not cover more than a 42-day period unless agreed upon.
- Assessment dates could not be within the first six weeks of the academic year.
- If assessment end dates were close to the end of the target population birth date period, NPMs were alerted not to conduct any make-up sessions beyond the date when the population births dates were valid.

Sampling Task 2: Stratification (and other information)

- Since explicit strata are formed to group similar schools together to reduce sampling variance and to ensure representativeness of students in various school types, using variables that might be related to outcomes, each participating country's choice of explicit stratification variables was assessed. If a country was known to have school tracking or distinct school programmes and these were not among the explicit stratification variables, a suggestion was made to include this type of variable.
- Dropping variables or reducing levels of stratification variables used in the past was discouraged and only accepted if the National Centre could provide strong reasons for doing so.
- Levels of variables and their codes were checked for completeness.
- If no implicit stratification variables were noted, suggestions were made about ones that might be used. In particular, if a country had schools with only male or only female students, and school gender composition was not among the implicit stratification variables, a suggestion was made to include this type of variable to ensure no sample gender imbalances. Similarly, if there were ISCED school level splits, the ISCED school level was also suggested as an implicit stratification variable.
- Without overlap control there is nearly as good control over the sample whether explicit or implicit strata are used. With overlap control some control is lost when using implicit strata, but not when using explicit strata. For countries which wanted overlap control with a national non-PISA survey, as many as possible of their implicit stratification variables were made explicit stratification variables.
- Checks were done to ensure there was only one student sampling option per explicit stratum.

Sampling Task 7a: National desired target population

- The total national number of 15-year olds of participating countries was compared with those from previous cycles. Differences, and any kind of trend, were queried.
- Large deviations between the total national number of 15-year-olds and the enrolled number of 15-year-olds were questioned.
- Large increases or decreases in enrolled population numbers compared to those from previous PISA cycles were queried, as were increasing or decreasing trends in population numbers since PISA 2000.

- Any population to be omitted from the international desired population was noted and discussed, especially if the percentage of 15-year-olds to be excluded was more than 0.5% or if it was substantially different or not noted for previous PISA cycles.
- Calculations did not have to be verified as in previous cycles as such data checks were built into the form.
- For any countries using a three-stage design, a Sampling Task 7a form also needed to be completed for the full national desired population as well as for the population in the sampled regions.
- For countries having adjudicated regions, a Sampling Task 7a form was needed for each region.
- Data sources and the year of the data were required. If websites were provided with an English page option, the submitted data was verified against those sources.

Sampling Task 7b: National defined target population

- The population value in the first question needed to correspond with the final population value on the form for Sampling Task 7a. This was accomplished through built-in data checks.
- Reasons for excluding schools for reasons other than special education needs were checked for appropriateness (i.e. some operational difficulty in assessing the school). In particular, school-level language exclusions were closely examined to check correspondence with what had been noted about language exclusions on Sampling Task 2.
- Exclusion types and extents were compared to those recorded for PISA 2009 and previous cycles. Differences were queried.
- The number and percentage of students to be excluded at the school level and whether the percentage was less than the guideline for maximum percentage allowed for such exclusions were checked.
- Reasonableness of assumptions about within-school exclusions was assessed by checking previous PISA coverage tables. If there was an estimate noted for “other”, the country was queried for reasonableness about what the “other” category represented. If it was known the country had schools where some of the students received instruction in minority languages not being tested, an estimate for the within-school exclusion category for “no materials available in the student’s language of instruction” was necessary.
- Form calculations were verified through built-in data checks, and the overall coverage figures were assessed.
- If it was noted that there was a desire to exclude schools with only one or two PISA-eligible students at the time of contact, then the school sampling frame was checked for the percentage of population that would be excluded. If a country had not exceeded the 2.5% school-exclusion guideline, excluding such schools was a possibility. Furthermore, if these schools would account for not more than 0.5% of students, and if within-school exclusions looked similar to the past and were within 2.5%, then the exclusion of these schools at the time of contact was agreed upon. There was one caveat for the agreement – that such exclusion not cause entire strata to be missing from the student data.
- The population figures on this form after school-level exclusions were compared against the aggregated school sampling frame enrolment. Differences were queried.
- For any countries using a three-stage design, a Sampling Task 7b form was also needed to be completed for the full national defined population as well as for the population in the sampled regions.
- For countries having adjudicated regions, a Sampling Task 7b form was needed for each region.
- Data sources and the year of the data were required. If websites were provided with an English page option, the submitted data was verified against those sources.

Sampling Task 8a: Sampling frame description

- Special attention was given to countries who reported on this form that a three-stage sampling design was to be implemented and additional information was sought from countries in such cases to ensure that the first-stage sampling was done adequately.
- The type of school-level enrolment estimate and the year of data availability were assessed for reasonableness.
- Countries were asked to provide information for each of various school types,[5] whether those schools were included on or excluded from the sampling frame, or the country did not have any of such schools. The information was matched to the different types of schools containing PISA students noted on Sampling Task 2. Any discrepancies were queried.
- Any school types noted as being excluded were verified as school-level exclusions on the Sampling Task 7b form. Any discrepancies were queried.

Sampling Task 8b: Sampling frame

- On the spreadsheet for school-level exclusions, the number of schools and the total enrolment figures, as well as the reasons for exclusion, were checked to ensure correspondence with values reported on the Sampling Task 7b form detailing school-level exclusions. It was verified that this list of excluded schools did not have any schools which only had one or two PISA-eligible students, as these schools were not to be excluded from the school sampling frame. Checks were done to ensure that excluded schools did not still appear on the other spreadsheet containing the school sampling frame.
- All units on the school sampling frame were confirmed to be those reported on the Sampling Task 2 as sampling frame units. The sampling unit frame number was compared to the corresponding frame for PISA 2009 as well as previous cycles. Differences were queried.
- NPMs were queried about whether or not they had included schools with grades 7 or 8, or in some cases those with grades 10 or higher, which could potentially have PISA-eligible students at the time of assessment even if the school currently did not have any.
- NPMs were queried about whether they had included vocational or apprenticeship schools, schools with only part-time students, international or foreign schools or schools not under the control of the Ministry of Education or any other irregular schools that could contain PISA-eligible students at the time of the assessment, even if such schools were not usually included in other national surveys.
- The frame was checked for all required variables: a national school identifier with no duplicated values, a variable containing the school enrolment of PISA-eligible students, and all the explicit and implicit stratification variables and all related levels as noted on Sampling Task 2, and that none had missing values.
- Any additional school sampling frame variables were assessed for usefulness. In some instances other variables were noted on the school frame that might also have been useful for stratification.
- The frame was checked for schools with only one or two PISA-eligible students. If no schools were found with extremely low counts, but the country's previous sampling frames had some, this was queried.
- The frame was checked for schools with zero enrolment. If there were none, this was assessed for reasonableness. If some existed, it was verified with the NPM that these schools could possibly have PISA-eligible students at the time of the assessment.

Sampling Task 9: Treatment of small schools and the sample allocation by explicit strata

- All explicit strata had to be accounted for on the form for Sampling Task 9.
- All explicit strata population entries were compared to those determined from the sampling frame.
- All small school analysis calculations were verified.
- It was verified that separate small school analyses were done for adjudicated or non-adjudicated oversampled regions (if these were different from explicit strata).
- Country specified sample sizes were monitored, and revised if necessary, to be sure minimum sample sizes for adjudicated regions were being met.
- The calculations for school allocation were checked to ensure that schools were allocated to explicit strata based on explicit stratum student percentages and not explicit stratum school percentages, that all explicit strata had at least two allocated schools, and that no explicit stratum had only one remaining non-sampled school.
- It was verified that the allocation matched the results of the explicit strata small school analyses, with allowances for random deviations in the numbers of very small, moderately small, and large schools to be sampled in each explicit stratum.
- The percentage of students in the sample for each explicit stratum had to be approximate to the percentage in the population for each stratum (except in the case of oversampling).
- The overall number of schools to be sampled was checked to ensure that at least 150 schools would be sampled.
- The overall number of students to be sampled was checked to ensure that at least 5 250 students would be sampled.
- Previous PISA response rates were reviewed and if deemed necessary, sample size increases were suggested.

Sampling Task 10: School sample selection

- All calculations were verified, including those needed for national study overlap control.
- Particular attention was paid to the required four decimal places for the sampling interval and the generated random number.

- The frame was checked for proper sorting according to the implicit stratification scheme, for enrolment values, and the proper assignment of the measure of size value, especially for very small and moderately small schools. The assignment of replacement schools and PISA identification numbers were checked to ensure that all rules established in the *Sampling Preparation Manual* were adhered to.

Sampling Task 11: Reviewing and agreeing to the Sampling Form

- The form for Sampling Task 11 was prepared as part of the sample selection process. After the PISA Consortium verified that all entries were correct, NPMs had one month to perform the same checks and to agree to the content in this form.

Sampling Task 12: School participation and data validity checks

- Extensive checks were completed on Sampling Task 12 data since it would inform the weighting process. Checks were done to ensure that school participation statuses were valid, that student participation statuses had been correctly assigned, and that all student sampling data required for weighting were available and correct for all student sampling options. Quality checks also highlighted schools having only one grade with PISA-eligible students, only one gender of PISA-eligible students, or schools which had noticeable differences in enrolled student counts than expected based on sampling frame enrolment information. Such situations were queried.
- Large differences in overall grade and gender distributions compared to unweighted 2009 data were queried.
- These data also provided initial unweighted school and student response rates. Any potential response rate issues were discussed with NPMs if it seemed likely that a non-response bias report might be needed.
- Large differences in response rates compared to PISA 2009 were queried.
- Participating countries doing CBA were expected to have data for CBA related variables. Any inconsistent or unexpected CBA data entries were queried.

Student samples

Student selection procedures in the Main Survey were the same as those used in the Field Trial. Student sampling was generally undertaken using the PISA Consortium software, *KeyQuest*, at the National Centres from lists of all PISA-eligible students in each school that had agreed to participate. These lists could have been prepared at national, regional, or local levels as data files, computer-generated listings, or by hand, depending on who had the most accurate information. Since it was important that the student sample be selected from accurate, complete lists, the lists needed to be prepared slightly in advance of the testing period and had to list all PISA-eligible students. It was suggested that the lists be received one to two months before the testing period so that the NPM would have adequate time to select the student samples.

Five countries (Chile, Germany, Iceland, Slovenia and Switzerland) chose student samples that included students aged 15 and/or enrolled in a specific grade (e.g. Grade 10). Thus, a larger overall sample, including 15-year-old students and students in the designated grade (who may or may not have been aged 15) was selected. The necessary steps in selecting larger samples are noted where appropriate in the following details:

- Iceland and Switzerland, for part of its grade sample only, used the standard method of direct student sampling described here.
- For Iceland, the sample constituted a de facto grade sample because nearly all of the students in the grade to be sampled were PISA-eligible 15-year-olds.
- Switzerland supplemented the standard sampling method with two list-based variations: one variation selected *TCS* students from just the PISA-eligible non-Grade 9 students followed by a census of Grade 9 students, and the other variation selected not more than *TCS* students from a list containing both PISA-eligible students and Grade 9 students.
- Germany supplemented the standard sampling method with an additional sample of grade-eligible students which was selected by first selecting Grade 9 classes within PISA sampled schools that had this grade.
- Slovenia used the direct student sampling method for two within-school samples, one for PISA-eligible students and a second sample for Grade 10 students that were not PISA-eligible.
- In Chile, the standard method was supplemented with additional grade-eligible students from a sample of Grade 10 classes within PISA sampled schools that had this grade.

Mexico selected a Grade 12 sample but accomplished this by having a completely separate sample of schools containing Grade 12 students.

Preparing a list of age-eligible students

Each school drawing an additional grade sample was to prepare a list of *age* and grade-eligible students that included all PISA-eligible students in the designated grade (e.g. Grade 10); and all other 15-year-old students (using the appropriate 12-month age span agreed upon for each participating country) currently enrolled in other grades. This form was referred to as a student listing form. The following were considered important:

- Age-eligible students were all students born in 1996 (or the appropriate 12-month age span agreed upon for the participating country).
- The list was to include students who might not be tested due to a disability or limited language proficiency.
- Students who could not be tested were to be excluded from the assessment after the student sample was selected. It was stressed that students were to be excluded after the students sample was drawn, not prior.
- It was suggested that schools retain a copy of the student list in case the NPM had to contact the school with questions.
- Student lists were to be up-to-date at the time of sampling rather than a list prepared at the beginning of the school year. Students were identified by their unique student identification numbers.

Selecting the student sample

Once NPMs received the list of PISA-eligible students from a school, the student sample was to be selected and the list of selected students (i.e. the student tracking form) returned to the school. NPMs were required to use *KeyQuest*, the PISA Consortium sampling software, to select the student samples unless otherwise agreed upon. For PISA 2012, all countries used *KeyQuest*.

Preparing instructions for excluding students

PISA was a timed assessment administered in the instructional language(s) of each participating country and designed to be as inclusive as possible. For students with limited assessment language(s) experience or with physical, mental, or emotional disabilities who could not participate, PISA developed instructions in cases of doubt about whether a selected student should be assessed. NPMs used the guidelines to develop any additional instructions; school co-ordinators and test administrators needed precise instructions for exclusions. The national operational definitions for within-school exclusions were to be clearly documented and submitted to the PISA Consortium for review before testing.

Sending the student tracking form to the School Co-ordinator and Test Administrator

The School Co-ordinator needed to know which students were sampled in order to notify students, parents and teachers to update information and to identify students to be excluded. The student tracking form was therefore sent approximately two weeks before the testing period. It was recommended that a copy of the tracking form be kept at the National Centre and the NPM send a copy of the form to the test administrator in case the school copy was misplaced before the assessment day. The Test Administrator and School Co-ordinator manuals (see Chapter 6) both assumed that each would have a copy.

In the interest of ensuring PISA was as inclusive as possible, student participation and reasons for exclusion were separately coded in the student tracking form. This allowed for students with Special Education Needs (SEN) to be included when their SEN was not severe enough to be a barrier to their participation. The participation status could therefore detail, for example, that a student participated and was not excluded for SEN reasons even though the student was noted with a special education need. Any student whose participation status indicated they were excluded for SEN reasons had to have an SEN code that explained the reason for exclusion. It was important that these criteria be followed strictly for the study to be comparable within and across participating countries. When in doubt, the student was included. The instructions for excluding students are provided in the PISA Technical Standards (see Annex F).

Definition of school

Although the definition of a "school" is difficult, PISA generally aims to sample whole schools as the first stage units of selection, rather than programmes or tracks or shifts within schools, so that the meaning of "between school variance" is more comparable across countries.

There are exceptions to this, such as when school shifts are actually more like separate schools than part of the same overall school. However, in some countries with school shifts this is not the case and therefore whole schools are used as the primary sampling unit. Similarly, many countries have schools with different tracks/programs but generally it is

recommended again that the school as a whole should be used as the primary sampling unit. There are some exceptions, such as the schools being split for sampling in previous PISA cycles (trends would be affected if the same practice was not continued), or if there is a good reason for doing so (such as to improve previously poor response rates, differential sampling of certain tracks or programmes is desired, etc.).

Sampling units to be used on school-level frames have been discussed with each country before the Field Trial. Figure 4.4 presents the comments from NPMs, in cases where "school" was not the unit of sampling. Where the Sampling Unit column indicates 'School', this means that the school was the sampling unit. Where it shows 'Other' then something else was used, as described in the comments. Figure 4.4 shows the extent to which countries do not select schools in PISA, but rather something else.

■ Figure 4.4 [Part 1/2] ■

Sampling frame unit

		Sampling unit school / Other	Sampling frame units comment
OECD	Australia	Other	Schools with more than one campus listed as separate entries
	Austria	Other	Either whole schools or programmes within schools
	Belgium	Other	A combination of whole schools (French- and German-speaking communities) and implantations (Tracks/programmes taught on a single address/location [administrative address]) (Flemish Community)
	Canada	School	
	Chile	Other	Whole schools except in the case of 80 schools, which were each split into two school parts, with one part treated as a school exclusion since it contains only excluded students. The other part corresponds to regular education programmes
	Czech Republic	Other	Basic school - whole school Special and practical school - whole school Gymnasium - pseudo schools according to the length of study (4 year gymnasium and 6 or 8 year gymnasium) Upper secondary vocational - pseudo schools (schools with *maturate*, schools without *maturate*)
	Denmark	School	
	Estonia	School	
	Finland	School	
	France	School	
	Germany	School	Exceptions in SEN schools
	Greece	School	
	Hungary	Other	Tracks in parts of schools on different settlements
	Iceland	School	
	Ireland	School	
	Israel	School	
	Italy	School	
	Japan	Other	Programme
	Korea	School	
	Luxembourg	School	
	Mexico	School	
	Netherlands	Other	Locations of (parts of) schools, often parts of a larger managerial unit
	New Zealand	School	
	Norway	School	
	Poland	School	
	Portugal	School	
	Slovak Republic	School	
	Slovenia	Other	Study program in ISCED3 schools and whole ISCED2 schools
	Spain	Other	Whole school is the option selected for Spain Only in the Basque Country (5% of Spanish population) the same school may be divided into three, each one corresponding to each linguistic model (A, B, D) within the region
	Sweden	School	
	Switzerland	School	
	Turkey	School	
	United Kingdom (excluding Scotland)	School	
	United Kingdom (Scotland)	School	
	United States	School	

■ Figure 4.4 [Part 2/2] ■

Sampling frame unit

		Sampling unit school / Other	Sampling frame units comment
Partners	Albania	School	
	Argentina	Other	Location of schools
	Brazil	School	
	Bulgaria	School	
	Colombia	School	
	Costa Rica	School	
	Croatia	Other	School locations
	Cyprus[1, 2]	School	
	Georgia	School	
	Hong Kong-China	School	
	Indonesia	School	
	Jordan	School	
	Kazakhstan	School	
	Latvia	School	
	Liechtenstein	School	
	Lithuania	School	
	Macao-China	School	
	Malaysia	School	
	Montenegro	School	
	Peru	School	
	Qatar	School	
	Romania	Other	School programmes
	Russian Federation	School	
	Serbia	School	
	Shanghai-China	School	
	Singapore	School	
	Chinese Taipei	School	
	Thailand	School	
	Tunisia	School	
	United Arab Emirates	Other	Separate curricula and also by gender. Whole schools sometimes
	Uruguay	School	
	Viet Nam	School	

1. Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Notes

1. Students were deemed participants if they gave at least one response to the cognitive assessment, or they responded to at least one student questionnaire item and either they or their parents provided the occupation of a parent or guardian (see Annex G).

2. Technical reference documents are available on the OECD PISA website: *www.oecd.org/pisa*.

3. Technical reference documents are available on the OECD PISA website: *www.oecd.org/pisa*.

4. Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

5. These include schools with multiple languages of mathematics instruction, vocational schools, technical schools, agriculture schools, and schools with only part-time students, schools with multiple shifts and so on.

References

Jaeger, R.M. (1984), *Sampling in Education and the Social Sciences*, Longman, New York.

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.

5

Translation and Verification of the Survey Material

One of the important aspects of quality assurance in PISA is to ensure that the instruments used in all participating countries to assess students' performance provide reliable and comparable information. In order to achieve this, PISA implemented strict procedures for adaptation, translation and verification of national versions of all survey instrumentation.

These procedures included:

- development of two source versions of the instruments, in English and French (except for the financial literacy and reading component skills options and for the operational manuals, provided only in English);
- double translation design;
- preparation of detailed instructions for the translation of the instruments for the Field Trial and for their review for the Main Survey;
- preparation of translation/adaptation guidelines;
- training of national staff in charge of the translation/adaptation of the instruments; and
- verification of the national versions by independent verifiers appointed by the Consortium.

DEVELOPMENT OF SOURCE VERSIONS

Part of the new test materials used in PISA 2012 was prepared by the Consortium test development teams on the basis of submissions received from the participating countries. Items were submitted by approximately 21 different countries, either in their national language or in English, for either or both of mathematics or problem solving. The other part of the material was prepared by the test development teams at: ACER, the University of Melbourne (both in Australia); aSPe (University of Liege, Belgium); DIPF (Deutschen Institut für Internationale Pädagogische Forschung), IPN (Leibniz-Institute for Science and Mathematics Education) and Heidelberg University (all three in Germany); NIER (the National Institute for Educational Policy Research, Japan); CRP-HT (Centre de Recherche Public – Henri Tudor, Luxembourg); ILS (the Department of Teacher Education and School Research, University of Oslo, Norway); and ETS (Educational Testing Service, United States). Then, all materials were circulated (in English) to the expert groups and the National Project Managers (NPMs) for comments and feedback.

The item development teams received specific information/training about how to anticipate potential translation and cultural issues. The document prepared for that purpose was mainly based on experience gained during previous PISA survey administrations. The item developers used it as a reference when developing and reviewing the items.

The French version was developed at an early stage through double translation and reconciliation of the English materials into French, so that any comments from the translation team could, along with the comments received from the expert groups and the NPMs, be used in the finalisation of both source versions.

Experience has shown that some issues do not become apparent until there is an attempt to translate the instruments. As in previous PISA survey administrations, the English to French translation process proved to be very effective in detecting residual errors overlooked by the test developers, and in anticipating potential problems for translation in other languages. In particular, a number of ambiguities or pitfall expressions could be spotted and avoided from the beginning by slightly modifying both the English and French source versions; the list of aspects requiring national adaptations could be refined; and further translation notes could be added as needed. In this respect, the development of the French source version served as a pilot translation, and contributed to providing NPMs with source material that was somewhat easier to translate and contained fewer potential translation problems than would have been the case had only one source been developed.

The final French source version was reviewed by a French domain expert, for appropriateness of the terminology, and by a native professional French proof-reader for linguistic correctness. In addition, an independent verification of the equivalence between the final English and French versions was performed by a senior staff member of cApStAn who is bilingual (English/French) and has expertise in the international verification of the PISA materials, and used the same procedures and verification checklists as for the verification of all other national versions.

Finally, analyses of all national versions adapted from the French source version were conducted, using the Field Trial item statistics from the five French-speaking countries and from all English testing countries participating in PISA 2012. The primary aim of these analyses was to increase the equivalence between the two source versions. National item

parameters in both source versions were aggregated and then compared in order to detect statistical differences. Further, within each source version, item parameters were compared to identify potential outliers that might be responsible for a significant difference at the aggregated level. Two main item parameters were analysed: *(i)* the item difficulty and *(ii)* the item discrimination.

The results were used during the revision of the French and English source versions for the Main Survey in three ways:

- For each statistically different item functioning in most French versions, careful word-by-word comparison with the English version was done. If a wording issue was identified, the French source version, and consequently all French national versions, were modified. This also sometimes resulted in a change in the English source in order to increase the equivalence between the sources.
- When statistically different item functioning occurred in only one or two French national versions, the versions that worked correctly were examined to see if national adaptations or wording improvements had been made by the National Centres. If this was the case, these were implemented in the French source version and consequently in all other national versions for the Main Survey.
- Items that worked well and similarly in English and French were kept unchanged.

DOUBLE TRANSLATION FROM TWO SOURCE LANGUAGES

Back translation has long been the most frequently used way to ensure linguistic equivalence of test instruments in international surveys. It requires translating the source version of the test (generally English language) into the national languages, then translating them back to English and comparing them with the source language to identify possible discrepancies.

A double translation design (i.e. two independent translations from the source language(s), and reconciliation by a third person) offers two significant advantages in comparison with the back translation design:

- Equivalence of the source and target versions is obtained by using three different people (two translators and a reconciler) who all work on both the source and the target versions. In a back translation design, by contrast, the first translator is the only one to simultaneously use the source and target versions.
- Discrepancies are recorded directly in the target language instead of in the source language, as would be the case in a back translation design.

PISA uses double translation from two different languages because both back translation and double translation designs fall short in that the equivalence of the various national versions depends exclusively on their consistency with a single source version (in general, English). In particular, one would wish the highest possible semantic equivalence (since the principle is to measure access that students from different countries would have to a same meaning, through written material presented in different languages). However, using a single reference language is likely to give undue importance to the formal characteristics of that language. If a single source language is used, its lexical and syntactic features, stylistic conventions and the typical patterns it uses to organise ideas within the sentence will have a greater impact on the target language versions than desirable (Grisay, 2003).

Some interesting findings in this respect were reported in the IEA[1]/reading comprehension survey (Thorndike, 1973), which showed a better item coherence (factorial structure of the tests, distribution of the discrimination coefficients) between English-speaking countries than across other participating countries.

Resorting to two different languages may, to a certain extent, reduce problems linked to the impact of cultural characteristics of a single source language. Admittedly, both languages used in PISA share an Indo-European origin, which may be regrettable in this particular case. However, they do represent relatively different sets of cultural traditions, and are both spoken in several countries with different geographic locations, traditions, social structures and cultures.

The use of two source languages in PISA results in other anticipated advantages such as the following:

- Many translation problems are due to idiosyncrasies: words, idioms, or syntactic structures in one language appear untranslatable into a target language. In many cases, the opportunity to consult the other source version may provide hints at solutions.

- The desirable or acceptable degree of translation freedom is very difficult to determine. A translation that is too faithful to the original version may appear awkward; if it is too free or too literary it is very likely to jeopardise equivalence. Having two source versions in different languages (for which the translation fidelity/freedom has been carefully calibrated and approved by Consortium experts including through analysis of Field Trial item statistics) provides national reconcilers with accurate benchmarks in this respect, which neither back translation nor double translation from a single language could provide.

The double translation and reconciliation procedure using both source languages was recommended in PISA 2012 as in previous survey administrations. Countries testing in non-Indo-European languages received additional support: the document "The challenge of translating PISA materials into non Indo-European languages", developed for PISA 2009, was re-circulated to countries for PISA 2012 as well.

PISA TRANSLATION AND ADAPTATION GUIDELINES

The *PISA Translation and Adaptation Guidelines*[2] had been extensively revised in 2009 with a view to obtaining a document that would be relevant to any PISA survey administration with relatively few administration-specific adjustments, which were made for 2012. The guidelines include:

- Instructions on double or single translation. Double translation (and reconciliation) is required for test and questionnaire materials, but not for manuals, coding guides and other logistic material. In double translation, it is recommended that one independent translator uses the English source version while the second uses the French version. In countries where the NPM has difficulty appointing competent translators from French/English, double translation from English/French only is considered acceptable; in such cases it is highly recommended to use the other source version for cross-checks during the reconciliation process insofar as possible.
- Instructions on recruitment and training.
- Security requirements.
- References to other documents, including technical guides for translating computer-based materials.

Other sections of the *PISA Translation and Adaptations Guidelines* are intended for use by the national translators and reconciler(s):

- Recommendations to avoid common translation traps.
- Instructions on how to adapt the test material to the national context.
- Instructions on how to translate and adapt questionnaires and manuals to the national context.
- The checklist used for the verification of PISA material.

As explained in the previous section, a separate document containing additional guidelines for translation into non Indo-European languages was also provided to countries.

After completion of the Field Trial, instructions on how to revise national version(s) were provided as a separate document known as *Material preparation Main Survey 2012*.[3]

TRANSLATION TRAINING SESSION

NPMs received sample materials to use when recruiting national translators and training them at the national level. The NPM meeting held in October 2010 in Budapest included a session on the Field Trial translation/adaptation activities in which recommended translation procedures, *PISA Translation and Adaptation Guidelines*, and the verification process were presented in detail.

At this meeting, countries were offered the opportunity to participate in a half day translation training workshop. Translators and NPMs attending the workshop received detailed information about the new PISA translation training "kit" designed to help National Centres implement PISA translation requirements in a more systematic way and were taken through presentations and hands-on exercises.

TESTING LANGUAGES AND TRANSLATION/ADAPTATION PROCEDURES

NPMs had to identify the testing languages according to instructions given in the *School Sampling Preparation Manual*[4] and to record them in a sampling form for agreement.

Prior to the Field Trial, NPMs had to fill in a translation plan (describing the procedures used to develop their national versions and the different processes used for translator/reconciler recruitment and training). Information about a possible national expert committee was also sought. This translation plan was reviewed by the Consortium for agreement.

Figure 5.1 summarises the Field Trial translation procedures for cognitive materials, as described in the confirmed country translation plans.

The distribution of the translation procedures used for the questionnaires was quite similar (with some shift) within and between countries. For example, the questionnaires in German were produced using the double translation from English and French sources while the cognitive units were produced using the double translation from English with cross-checks against the French version.

There is a domain effect in the procedure changes compared to PISA 2009. Some countries (e.g. Germany and Norway) that used double translation from both English and French sources in 2009 chose double translation from English source with cross-checks against the French source version in 2012 because they could not find translators from French with good experience in the mathematics domain.

■ Figure 5.1 ■

Field Trial 2012 translation procedures

Activity	Paper-based instruments	Computer-based instruments*
Double translation from English and French source versions	18	14
Double translation from English source version with cross-checks against the French source version	10	7
Double translation from French source version with cross-checks against the English source version	1	0
Double translation from English source version only	20	7
Adaptations in one of the source versions	18	15
Adaptations made in a borrowed verified version	13	16
Adaptations made in a verified common version	4	4

*Note that some countries did not participate in any of the computer-based assessments. Note also that in some cases, countries such as Germany and Belgium (Flemish Community) borrowed part of the computer-based instruments (problem solving) but produced their own computer-based mathematics instruments because the country from which they borrowed the problem-solving version did not participate in the computer-based assessment of reading literacy option.

Countries sharing a testing language were strongly encouraged to develop a common version in which national adaptations would be inserted or, in the case of minority languages, to borrow an existing verified version. There is evidence from all previous survey administrations (PISA 2000 through PISA 2009) that high quality translations and high levels of equivalence in the functioning of items was best achieved in the three groups of countries that shared a common language of instruction (English, French and German) and could develop their national versions by introducing a limited number of national adaptations in the common version. Additionally, a common version for different countries sharing the same testing language implies that all students instructed in a given language receive booklets that are as similar as possible, which reduces cross-country differences due to translation effects.

Co-operation between countries sharing a same language was hence fostered and facilitated by the Consortium: workable models were designed so that verified versions from one country could be adapted by a second country.

- In the case of the Chinese-speaking participants Hong Kong, Macao, Shanghai and Chinese Taipei, a common base version was developed jointly, verified centrally, and then adapted for local usage by each of the four entities. Each national version was then verified separately.
- The model followed by German-speaking countries was (again) highly efficient: the German version of each of the components of the assessment material was double translated and reconciled by one of the countries, then verified, then adapted by the other countries who administered that component. The adapted versions were then verified.
- Spain and Mexico were the Spanish-speaking countries with the earliest testing window, and they developed their national versions following the standard PISA procedures. Their national versions were verified by cApStAn and then, once Mexico and Spain had accepted/rejected verifier interventions as needed, "clean" verified versions were made available to the other Spanish-speaking countries. Costa Rica used the verified Mexican version; Argentina, Chile and Uruguay used the verified Spanish version from Spain to prepare their national version. Colombia and Peru developed their own versions and used the verified version from Spain for cross-checks.

- Other co-operation models involving the use of borrowed versions included: Greece and Cyprus[5] for Link items; Estonia and Latvia for the Russian version of the financial literacy option; the Russian Federation made its materials available to Estonia and Latvia; Italy and the Canton Ticino in Switzerland each translated from one source language and then held a meeting to reconcile and implement national adaptations; Montenegro made its version available to Serbia; and the United Arab Emirates borrowed the verified Jordanian version of paper-based Mathematics units.

TESTING LANGUAGES

A total of 98 national versions of the materials were used in the PISA 2012 Main Survey, in 46 languages. The languages were: Albanian, Arabic, Basque, Bulgarian, Chinese (traditional script), Chinese (simplified script), Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Irish, Italian, Japanese, Kazakh, Korean, Latvian, Lithuanian, Malaysian, Norwegian (Bokmål and Nynorsk), Polish, Portuguese, Romanian, Russian, Serb (Ekavian and Yekavian variants), Slovak, Slovenian, Spanish, Swedish, Thai, Turkish, Valencian, Vietnamese and Welsh.

International verification (described in section below) occurred for 85 national versions out of the 98 used in the Main Survey.

International verification was not implemented when a testing language was used for minorities that make less than 10% of the target population or when countries borrowed a version that had been verified at the national level without making any adaptations. This concerned 18 versions across the following countries: Belgium (German), Finland (Swedish), Hong Kong-China (English), Ireland (Irish), Italy (Slovene and German), Liechtenstein (German), Lithuania (Polish and Russian), Macao-China (Portuguese), Montenegro (Albanian), Romania (Hungarian), Serbia (Hungarian), the Slovak Republic (Hungarian) – for which however financial literacy was verified, Slovenia (Italian), Spain (Valencian), Sweden (English) and the United Kingdom (Welsh).

Note that among these 18 versions, only three (Irish, Valencian and Welsh) were only verified at the national level. All other versions were prepared using internationally verified versions.

INTERNATIONAL VERIFICATION OF THE NATIONAL VERSIONS

As in previous PISA survey administrations, one of the most important quality control procedures implemented to ensure high quality standards in the translated assessment materials for PISA 2012 was to have an independent team of expert verifiers, appointed and trained by the Consortium; verify each national version against the English and/or French source versions.

The Consortium undertook international verification of all national versions in languages used in schools attended by more than 10% of the country's target population. For a few minority languages, national versions were only developed (and verified) in the Main Survey phase.

The main criteria used to recruit verifiers of the various national versions were that they had:

- native command of the target language;
- professional experience as translators from English or French or from both English and French into their target language;
- as far as possible, sufficient command of the second source language (either English or French) to be able to use it for cross-checks in the verification of the material (note that not all verifiers are proficient in French, but this is mitigated by the fact that the cApStAn reviewer and the translation referee have command of French);
- as far as possible, familiarity with the main domain assessed (in this case, maths);
- a good level of computer literacy; and
- as far as possible, experience as teachers and/or higher education degrees in psychology, sociology or education.

Verifier training sessions were held prior to the verification of both the Field Trial and the Main Survey materials. Attendees received copies of the PISA information brochure, *Translation Guidelines*, the English and French source versions of the material and a *Verification Checklist*. The training sessions focused on:

- presenting verifiers with PISA objectives and structure;

- familiarising them with the material to be verified, the verification procedures, and the software tools to be used (in particular, the Open Language Tool (OLT) software used for computer-based materials);
- reviewing and extensively discussing the *Translation Guidelines* and the *Verification Checklist*;
- conducting hands-on exercises on specially “doctored” target versions;
- arranging schedules and dispatch logistics; and
- security requirements.

Verification procedures have been continually improved throughout each PISA round, based on the experience and learning from previous rounds. In the following subsections we review the “state of the art” of procedures for the different components subject to verification. These included in 2012: the (first phase) verification of new test units (both paper-based and computer-based), of the “booklet shell” (paper-based), of practice units/widgets (computer-based), of questionnaires (including an optional online School Questionnaire); the “convergence check” of link units (both paper-based and computer-based); the final check of assembled test booklets, questionnaire booklets, computer-based test units, and coding guides; and the verification of (selected parts of) operational manuals.

Verification of paper-based test units

Since the PISA 2000 Main Survey, verifiers enter their suggested edits in Microsoft Word® files (mostly in item pool format, i.e. including coding sections) using the track changes mode. This facilitates the revision of verified materials by the National Centre, which can directly “accept” or “refuse” the edits proposed.

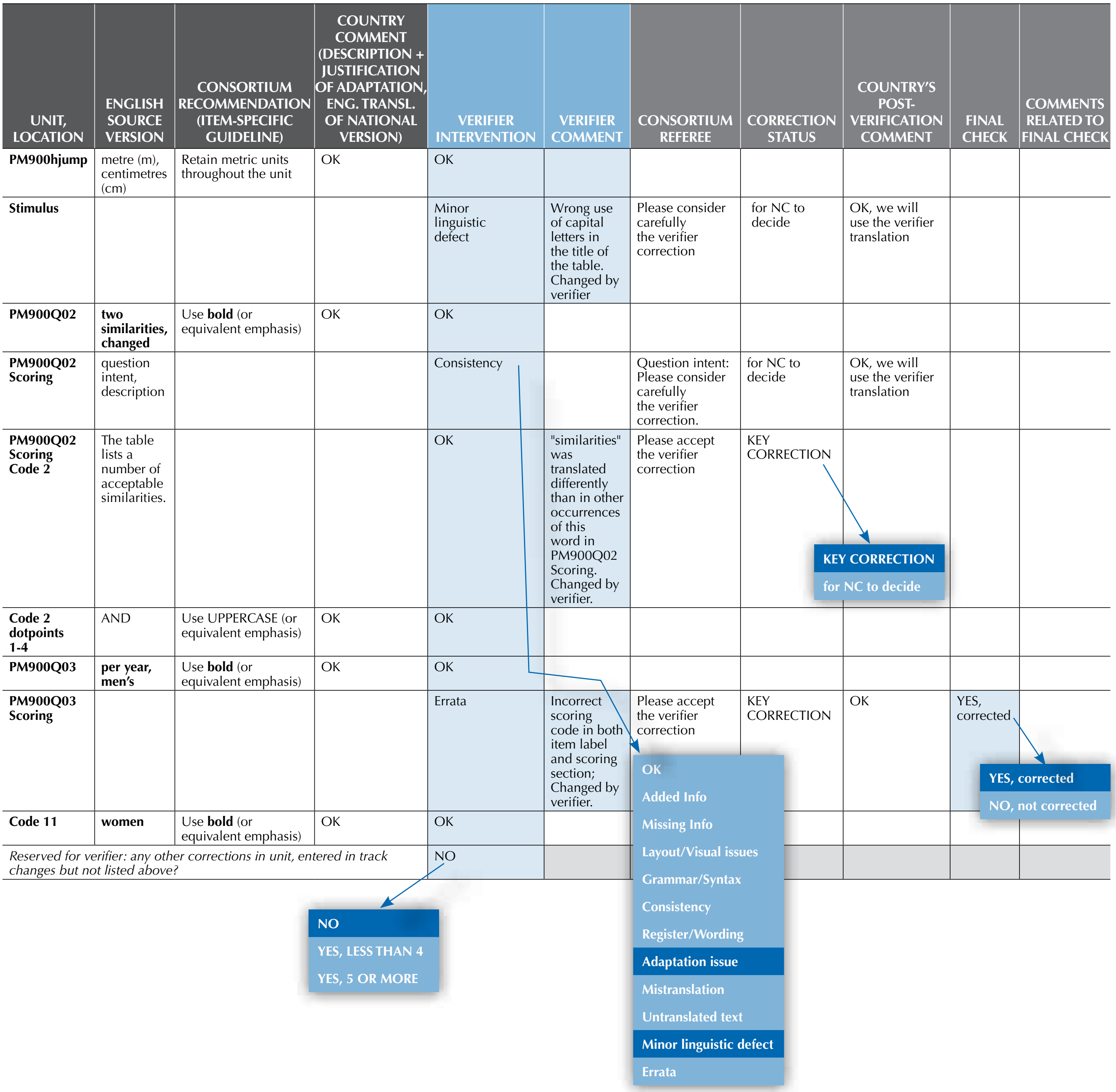
Since the PISA 2003 Main Survey, the mainstay of the verification procedure for test units has been the Test Adaptation Spreadsheet (TAS). Figure 5.2 shows a sample Test Adaptation Spreadsheet from the PISA 2012 Field Trial. The aim of this form is to function as an aid to translators, reconcilers, and verifiers (through the increasing use of item-specific translation/adaptation guidelines); as a centralised record of national adaptations, of verifier corrections and suggestions; as a way of initiating discussions between the National Centre and the Consortium referee; as a way of recording the implementation status of “key corrections” in test booklets; and as a tool permitting quantitative analysis of verification outcomes.

Some points of note are:

- Since PISA 2003, and increasingly so with each round, the column “Consortium Recommendation” is used to list item-specific translation/adaptation guidelines. These complement the general translation/adaptation guidelines and the translation notes embedded in source unit files with additional advice covering recommended, allowed or proscribed adaptations, literal or synonymous matches to be maintained, other psychometric characteristics to be considered (e.g. relative length or other patterns in multiple choice responses), desirable register of terms to be maintained, emphasis to be echoed, tips for the translation of difficult or idiomatic terms, etc. The verification co-ordinators consider that the generalised use of item-specific guidelines (for the attention of both translators and verifiers) is a significant breakthrough for translation quality assurance and quality control.
- Since PISA 2006, verifiers are instructed to document their “significant” verification interventions in the test adaptation spreadsheet, with a view to formalising the process by which a) the Consortium translation referees are informed of important issues and can liaise, when needed, with the test developers; b) if there is disagreement with the National Centre, a back-and-forth discussion ensues until the issue is resolved; c) key corrections in test materials are pinpointed so that their implementation can be checked at the Final Optical Check (FOC) phase. In the PISA 2000 and PISA 2003 rounds, this process was less structured.
- As of the PISA 2009 Main Survey, a conceptual change was introduced with regard to defining “significant” verification interventions tracked in the test adaptation spreadsheet. It was deemed desirable to reduce variability in the choice that verifiers make whether to report an intervention in the test adaptation spreadsheet or only correct in track changes in the unit, and to ensure that all potentially serious corrections are included in the test adaptation spreadsheet. This so that they may acquire “key correction” status and be checked during the Final Optical Check (FOC). The criterion was thus revised from “distinguish between purely linguistic issues and those that may potentially threaten equivalence” (used formerly) to “distinguish between minor errors or suggestions for improvement (that would not really affect the instruments if not corrected) and serious or potentially serious errors that require action.”

■ Figure 5.2 ■

Sample Field Trial Test Adaptation Spreadsheet (TAS) for a new PISA 2012 mathematics unit



UNIT, LOCATION	ENGLISH SOURCE VERSION	CONSORTIUM RECOMMENDATION (ITEM-SPECIFIC GUIDELINE)	COUNTRY COMMENT (DESCRIPTION + JUSTIFICATION OF ADAPTATION, ENG. TRANSL. OF NATIONAL VERSION)	VERIFIER INTERVENTION	VERIFIER COMMENT	CONSORTIUM REFEREE	CORRECTION STATUS	COUNTRY'S POST-VERIFICATION COMMENT	FINAL CHECK	COMMENTS RELATED TO FINAL CHECK
PM900hjump	metre (m), centimetres (cm)	Retain metric units throughout the unit	OK	OK						
Stimulus				Minor linguistic defect	Wrong use of capital letters in the title of the table. Changed by verifier	Please consider carefully the verifier correction	for NC to decide	OK, we will use the verifier translation		
PM900Q02	**two similarities, changed**	Use **bold** (or equivalent emphasis)	OK	OK						
PM900Q02 Scoring	question intent, description			Consistency		Question intent: Please consider carefully the verifier correction.	for NC to decide	OK, we will use the verifier translation		
PM900Q02 Scoring Code 2	The table lists a number of acceptable similarities.			OK	"similarities" was translated differently than in other occurrences of this word in PM900Q02 Scoring. Changed by verifier.	Please accept the verifier correction	KEY CORRECTION			
Code 2 dotpoints 1-4	AND	Use UPPERCASE (or equivalent emphasis)	OK	OK						
PM900Q03	**per year, men's**	Use **bold** (or equivalent emphasis)	OK	OK						
PM900Q03 Scoring				Errata	Incorrect scoring code in both item label and scoring section; Changed by verifier.	Please accept the verifier correction	KEY CORRECTION	OK	YES, corrected	
Code 11	**women**	Use **bold** (or equivalent emphasis)	OK	OK						
Reserved for verifier: any other corrections in unit, entered in track changes but not listed above?				NO						

- Since the PISA 2006 Main Survey, an innovation in the test adaptation spreadsheet is that verifiers use a scroll-down menu to categorise issues. As before, an additional comments field allows verifiers to explain their intervention with a back translation or description of the problem. The purpose of the categorisation is to reduce variability in the way verifiers document their verification; to make it easier for the Consortium referee to judge the nature of an issue and take action as needed; and to provide an instrument to help assess both the initial quality of national versions and the quality of verifiers' output.
- In the PISA 2012 Field Trial verification, there were 12 intervention categories for the verifier to choose from. Four new categories were introduced since PISA 2009: "OK", "Minor linguistic defect", "Untranslated text" and "Errata". The "OK" category implies a formal commitment from the verifier: s/he confirms having checked linguistic correctness and equivalence of the text element/segment and, if applicable, its conformity to an item-specific translation/adaptation

guideline. The category “Minor linguistic defect” was intended for minor issues that the verifier may wish to mention. “Untranslated text” was included as a separate category for cases in which part of the text remained in source language, for example in graph captions. “Errata” was introduced to differentiate between errors originating from the translation process and errors originating from defects in the source version.

- In training verifiers, special attention has been given since PISA 2009 to harmonising comment-writing practices. The life cycle of a comment makes it necessary to express it in such a way that it will be useful for both the National Centre and for the Consortium referee. Furthermore, the Final Optical Check (FOC) reviewer, who is not always the same person as the verifier, must be able to verify at final check whether a correction has been implemented or not. The following guidelines were set for verifier comments:
  - comments should be understandable to the Consortium referee who does not always understand the target language and preferably only looks at the test adaptation spreadsheet and the source version when reviewing comments;
  - specify in what way the target deviates from the source (rather than giving instructions on how to fix it, quoting the source, or explaining how the text has been corrected);
  - mention whether the verifier has made the correction or not for and why (e.g. because the verifier is unable to do it, or is not sure how to go about it);
  - comments should be factual and written in a clear and unemotional way and opinion statements should be kept to a minimum; and
  - each comment should relate to the category label selected.

For the PISA 2012 Field Trial verification, all translated versions were submitted for verification in item pool format (one unit per Word® file, including related coding sections), while a majority of national versions adapted from the English or French source versions were submitted in booklet format, i.e. as assembled test booklets (without coding sections). A special TAS was created to accommodate this, and later customised TAS were produced for the separate verification of coding guides. Customised TAS were also developed for other “homolingual” national versions adapted from a previously verified version in the same language.

Verifiers charged with homolingual versions were given modified instructions, with a focus on checking whether the adaptations made by the National Centre are appropriate and do not introduce errors or inconsistencies, checking for possible undocumented changes, and checking whether the National Centre has missed making any needed or recommendable adaptations to suit local context.

Verifiers for respectively English and French versions (and the cApStAn reviewer charged with co-ordinating and reviewing their work before delivery) were further instructed to consult a list of “tips”, including spelling and other adaptable elements but especially errata, errata candidates and “quasi-errata” (suggestions for improving the source) built up with each successive verification.

Main Survey verification of paper-based test units

Main Survey verification is, in essence, a second verification of materials already verified once before the Field Trial. In PISA 2009, Main Survey materials were fully (re-)verified, while in PISA 2012, the Main Survey units mostly underwent a *focused* verification: the verifiers concentrated on what had been or should have been changed since the Field Trial.

For paper-based assessment items, the assignment specification of verifiers included *(i)* checking correct implementation of Field Trial to Main Survey Consortium changes; *(ii)* assessing whether changes proposed by countries are acceptable by PISA standards; *(iii)* verifying whether those changes were correctly and consistently implemented; *(iv)* addressing issues from dodgy item reports; *(v)* checking whether the country used its final Field Trial version to prepare its Main Survey version; *(vi)* carrying out a selective full (re-)verification of earmarked units and, in some problem cases, of the entire national version. We refer to steps *(i)* to *(iv)* as “focused” verification, as opposed to “full” verification (sentence-by-sentence verification of national vs. source version on the entire materials), which is the norm for Field Trial verification.

For computer-based materials, as we shall see later, it was possible to perform an entirely “safe” focused verification i.e. there was no need for selective re-verification of parts that were effectively guaranteed to be final Field Trial versions in which no Field Trial to Main Survey changes were made or needed. Technological solutions implemented in the Translation Management System (TMS) allowed to a) ensure that countries would start out from correct “base” national versions; and b) mark segments that were (or should have been) revised by countries.

Paper-based materials were submitted for verification in cluster format, with the exception of a small group of early-testing countries which submitted units for verification in item pool format. The TAS were organised by cluster. The purpose of these spreadsheets was to list all Consortium changes; to provide countries with a file in which to list the changes they would want to introduce in their Main Survey version to address issues detected through the analysis of Field Trial results; to document verifier interventions and the follow-up of these interventions as indicated by the Consortium's translation referees.

■ Figure 5.3 ■

Field Trial to Main Survey Consortium changes documented in the TAS

UNIT, LOCATION	CLUSTER	Field Trial (FT) > Main Survey (MS) CHANGES			COUNTRY COMMENT (Description and/or justification of change, plus English translation of new national version)
		ENGLISH SOURCE VERSION (in case of national change)	NATIONAL FT VERSION (or English FT source version)	NATIONAL MS VERSION (or English MS source version)	
PM967Q03	**PM7B-12**	1st response option in table	α must be an even number	α must be an even number of degrees	Consortium change (addition of 'degrees')
		2nd response option in table	360° divided by α must be a whole number. Yes / No	~~360° divided by α must be a whole number.~~ ~~Yes / No~~	Consortium change (deletion of 2nd response option)

The TAS were dispatched to the National Centres together with the final source versions, with rows pertaining to scoring sections greyed out (indicating that scoring sections would not be verified at the same time as items, see later).

Countries were asked to document in the TAS any Field Trial to Main Survey changes they wished to make in their units. Such changes were expected to be relatively rare, since the instructions to the National Centres were to *"refrain from over-revisions just aimed at finding more "elegant", "literary" or "efficient" words and syntactical structures. Such preferential or cosmetic modifications might affect the item difficulty in unexpected ways, perhaps introducing flaws in items that had no problem in the FT."* Verifiers were instructed to gauge national changes in light also of the above consideration. For each national Field Trial to Main Survey change, the National Centres were asked to enter the Field Trial version, the suggested revised Main Survey version, and to explain in English the rationale behind the change.

■ Figure 5.4 ■

Country change documented in the TAS

Field Trial (FT) > Main Survey (MS) CHANGES			COUNTRY COMMENT (Description and/or justification of change, plus English translation of new national version)
ENGLISH SOURCE VERSION (in case of national change)	NATIONAL FT VERSION (or English FT source version)	NATIONAL MS VERSION (or English MS source version)	
Statement 2: In December 2010, Locality 1 had more than half of its total population employed.	Τον Δεκέμβριο του 2010, περισσότεροι από το μισό πληθυσμό του δήμου 1 ήταν εργαζόμενοι.	Τον Δεκέμβριο του 2010, πάνω από το μισό του συνολικού πληθυσμού του Δήμου 1 ήταν εργαζόμενοι.	Slightly rephrased and addition of the word «total», missing in FT

Prior to the start of the verification, cApStAn received selected information from the Field Trial item analysis. This mainly concerned items for which countries suspected that a translation or adaptation issue might have affected response patterns. So, items were singled out whenever they showed differential item functioning in the Field Trial and either the country or the verifier felt that a national Field Trial to Main Survey change might address the problem. Information about such "dodgy" items was incorporated in the TAS of the affected versions prior to sending the materials to the verifier.

For those language versions that were deemed to be of unsatisfactory overall quality during the Field Trial verification, a full verification of all units was performed before the Main Survey. This was the case for three national versions.

Although there were random checks (six units were earmarked for full verification in each version), there was no systematic check whether National Centres had used the final Field Trial versions as a basis when implementing Field Trial to Main Survey changes, and no systematic check whether countries had made undocumented edits in addition to documented edits. However, verifier interventions in Main Survey units revealed this type of problem in some language versions: one country implemented Field Trial to Main Survey changes in unverified financial literacy units, and the same was observed in Mathematics clusters from two countries. Some countries testing in English submitted Main Survey source versions for verification (i.e. ignoring all national adaptations made during the Field Trial verification). Appropriate corrective action (e.g. re-implementation of key corrections or national adaptations from the Field Trial stage) was taken in such cases.

Verification of the booklet shell

Since PISA 2006, the booklet shell has been handled as a separate component subject to verification. In PISA 2012, the booklet shell included the *Booklet Cover*, the *Formula Page*, the *General Directions*, sources/copyright acknowledgements, and the *Calculator Use and Effort Survey*. For countries taking the financial literacy option, the latter was replaced by two different forms of a short *"Questions about your experiences on money matters"* survey in the financial literacy booklets. It was dispatched together with a booklet adaptation spreadsheet which has the same structure as the test adaptation spreadsheet, and is verified following the same procedure as the test units.

Verification of computer-based test units

Of the 66 countries in PISA 2012 Field Trial, 43 participated in the computer-based assessment (CBA) taking at least the problem solving core component. Of these 43 countries, 29 countries participated also in the optional computer-based mathematics assessment, and 14 countries participated also in the optional digital reading assessment. Nineteen countries had participated to the digital reading assessment (DRA) within the framework of the PISA 2009 Main Survey. All other countries were "new" in administering computer-based materials.

Computer-based units were translated and verified using the Open Language Tool (OLT) software on XLIFF (tagged XML Localisation Interchange File Format) files which were exchanged, previewed and archived on the ACER Translation Management System (TMS), a web-based upload-download platform.

To perform the verification task, verifiers used a version of the OLT software especially customised for PISA by DIPF. They were instructed to verify the text segments one by one, comparing target version appearing on the right side of the OLT interface to source version appearing on the left side, consulting previews and the comment window to see item-specific guidelines and National Centre comments. They made corrections as needed, documenting their interventions in the comment window within OLT, including selection of the appropriate intervention category using a drop-down menu. The basic verification task, the checklist and the intervention categories were the same as for the paper-based materials. However, there was no TAS (replaced by the comment field carried within each XLIFF file and later in the process by the differences report). Also, as there is no "track changes" facility in the OLT, verifiers used the segment status marker (a feature of OLT) to differentiate edited versus unchanged segments.

Once a domain was verified, reviewed and "finalised" on the TMS, the Consortium Referee obtained an automatically generated "differences report" in Excel® format from the TMS. This report was used as the TAS for the paper-based units for a key corrections process (but using the international contractor's website MyPISA, *http://mypisa.acer.edu.au*, for various phases).[6] Key corrections were negotiated between Consortium referee and National Centre, the National Centre uploaded revised XLIFF files on the TMS for final check, the final check reviewer checked the correct implementation of key corrections and either released the files to ACER for national version construction or released them back to the National Centre for last-minute corrections.

Arrangements for the verification of homolingual versions – computer-based units adapted from the English or French source versions, from a common version (e.g. Chinese) or from a borrowed verified same-language version (e.g. Chile borrowing from Spain) were similar to those for paper-based cognitive units. In particular, English and French versions benefited from a co-ordination process similar to the one implemented for cognitive materials: a list of "tips" for verification of computer-based units, including spelling and other possibly adaptable elements, and especially errata, errata candidates and "quasi-errata" (suggestions for improving the source) was maintained, built up, and used in each successive verification.

As already mentioned, in the case of computer-based units a breakthrough was achieved in PISA 2012 regarding a "safe" exclusively focused verification at Main Survey phase. Different from paper-based materials, the process for CBA ensures that countries started out from their verified and finalised Field Trial versions, and prevents undocumented Field Trial to Main Survey changes that may escape verification. Further, the segment status used for marking unchanged segments ("Approved") also "lightly locks" such segments: both the National Centres and the verifiers could not change these segments by accident. This helps both in discouraging "cosmetic" changes by National Centres and verifiers (they have to unlock "Approved" segments where they purposefully choose to make a change), and in easing the post-verification review by National Centres (at that stage, they can ignore the "Approved" segments in verified XLIFF files).

Verification (convergence check) of coding sections/coding guides

For the Field Trial, coding sections were verified either together with stimuli/items – in Word® files submitted in item pool format – or separately, the latter being the norm for versions adapted from the French or English source versions and for

the computer-based (CB) coding guides (problem solving, mathematics, reading). The process was the same as for other paper-based (PB) test materials, but there was no final check of key corrections related to coding sections at Field Trial stage. It was the countries' responsibility to implement key corrections in the coding guides and to echo edits that arose from the coder training meeting in Rome in February 2011, which were reflected in the final "post-Rome" version of the coding guides released in early March 2011. This meant differential treatment between countries that submit their units for verification earlier or later: early-testing countries (other than those testing in English or French) had their coding sections verified before the final source version of the coding guides was released.

For the Main Survey, consolidated PB and CB Coding Guides were verified (or "convergence-checked", see next paragraph) separately from assessment items, after the source version of coding guides was finalised in March 2012, incorporating late edits arising from the coder training meeting in Salzburg. Introductions were fully verified. For the coding sections, the procedure was the same as for the clusters: countries document in the TAS any national FT (Field Trial)>MS (Main Survey) changes they want to make, and verifiers are asked to check that such changes are acceptable and correctly and consistently implemented in the guide, in addition to checking correct implementation of Consortium FT>MS changes. Because in the Field Trial the coding guides were verified prior to the coder meeting for most language versions and post-coder training coding guides had not been subject to verification, checking the correct implementation of a sample of late FT changes was included in the Main Survey verification procedure.

Verification (convergence check) of link units

Link units were verified for three new versions from Cyprus[7] (Greek and English) and Viet Nam, which had not participated in PISA 2009. Cyprus[8] adapted their Greek units from the version used by Greece in 2009 and their English units from the English source version, while Viet Nam provided new translations of all items. For these countries, link units were verified following the same procedure as the new mathematics or financial literacy units (full verification at Field Trial, focused verification at Main Survey).

For countries and economies that had already participated in PISA, the Main Survey process included a special focus on checking identicalness between national versions of link items (which were not field-trialled) versus the versions used previously. This was assessed by means of a semi-automated convergence check, whereby the versions the countries submitted in 2012 were compared to the final PISA 2009 Main Survey versions that countries had uploaded to MyPISA in 2009. It remains a concern that a vast majority of countries either failed to use the final 2009 versions to assemble the 2012 versions or introduced a number of changes without documenting them, often dismissing these modifications as unimportant – whereas the literature indicates that even minute changes can potentially affect response patterns.

A convergence check was organised to compare the content of the nine link clusters (three for science, three for mathematics and three for reading) submitted by countries for the PISA 2012 Main Survey with the content of the corresponding clusters administered in the PISA 2009 Main Survey. To avoid jeopardising the trend, no changes were allowed, except if it seemed indispensable to correct an outright error. For such exceptions, National Centres had to describe the rationale of each edit in a special "link TAS" prior to submission of the link clusters, which was made available only to those countries that requested it. Each of these edits was first commented by the verifier, then by the Consortium referee, and if needed by the test developers. Each change request was ultimately approved or rejected, following a negotiation phase with the National Centre if needed.

A selection of 5 science link units (out of 18), 7 mathematics link units (out of 25) and 5 reading literacy link units (out of 13) was earmarked for a comprehensive convergence check. The selection of units was not arbitrary: *(i)* based on TAS statistics per unit from the PISA 2009 Main Survey, units with the highest number of verifier interventions in 2009 were selected; *(ii)* units from PISA 2009 that underwent last minute changes were also included. The National Centres did not know which units were selected for convergence check.

The convergence check was performed using a combination of an automated PDF® comparison functionality and a manual check. The verifier documented all discrepancies in the link TAS. The verifier was (in a first stage) not asked to assess the significance of identified changes. If it became apparent that the country had made a great number of changes, the convergence check was interrupted and the National Centre was asked *(i)* to check whether the correct version of link units had been used to prepare the PISA 2012 Main Survey link clusters; *(ii)* if yes, to either make formal, well-documented requests for these changes; or *(iii)* to revert back to the 2009 version. In case the National Centre insisted on making changes, the full negotiation procedure was launched: the link TAS was submitted with a description of the rationale for each change; the verifier commented on them, then the Consortium referee, and then the test

developers. The TAS was then sent back to the country, and the country was asked to re-submit link clusters, in which *only* those changes that had been approved should be implemented. In this second convergence check, at least one (randomly selected) unit per domain was added to the convergence check.

For 11 out of 77 language versions, the convergence check did not reveal any changes from PISA 2009 to PISA 2012. For 10 language versions, changes were requested *before* the convergence check. The link clusters submitted for the convergence check were rejected for 11 language versions due to a high number of undocumented changes. In the case of Kazakhstan (both Russian and Kazakh versions) too many changes were found also in the second convergence check. As a result, these two versions were fully verified. In a majority of language versions, 1 to 10 discrepancies were identified in the units selected for convergence check. All discrepancies were regarded as "key corrections", meaning that compliance with the Consortium's recommendation for each discrepancy documented in the TAS is checked at final check stage.

The convergence check procedure of the link units revealed two major weaknesses: firstly the National Centres often failed to comply with the requirement that the link units used in 2012 should be fully identical to the units used in 2009. Compliance with the procedure was achieved only by a minority of National Centres. Secondly; it seems that many National Centres found it difficult to locate the correct, final versions of their 2009 units. This may be due to changes in national teams since 2009 or to defective version management policies.

The scoring sections of the new mathematics units were verified separately in the form of consolidated coding guides; therefore also the convergence check of the scoring sections of link units took place separately from the convergence check of clusters. The procedure was the same as for clusters: the same "link TAS" was used and the selection of units subject to this check was the same. Very few countries documented the changes they wished to make, and most countries made some changes. As the scoring sections are less sensitive in terms of impact on maintaining the link, minor changes (punctuation issues, typos corrected, and changes in the labelling of scoring rubrics made for consistency across domains or with new units) were mostly approved by the Consortium referee, even if the country had not specifically requested permission to make such changes. At final check of coding guides it was checked if residual issues in the coding sections of the link units were correctly addressed.

Verification of questionnaires

Questionnaires are submitted for verification together with an agreed Questionnaire Adaptation Spreadsheet (QAS). The purpose of the QAS is to document all content-related deviations from the international reference versions. Such national adaptations are subject to clearance by the questionnaire team before the material is submitted for verification.

The verifiers' brief (successively refined throughout PISA survey administrations) is now defined as checking whether target questionnaires are linguistically correct and faithful to either the source version (when no adaptation is made) or the approved English translation of the national version (when an adaptation is made). With a view to this, verifiers are instructed:

- to check whether the back translation of the agreed adaptation is faithful;
- to check whether the agreed adaptation is correctly reflected in the questionnaire;
- to check the questionnaires for undocumented adaptations (deviations from the source not listed in the QAS) and report them; and
- to check linguistic correctness (grammar, spelling, etc.) of the entire target version.

In the same manner as for test units, corrections are entered in the questionnaires using the track changes mode (exception: the optional online school questionnaires, for which a modified procedure was used to track corrections), while verifier comments are entered in the verifier column of the QAS.

In essence the Field Trial verification procedure for questionnaires did not change significantly since PISA 2009. The main changes consisted in *(i)* using MyPISA for each milestone; *(ii)* introducing verifier intervention categories in questionnaire verification; *(iii)* having a structured key corrections check, similar to the final check process for booklets.

Although there were no special "homolingual" procedures for the verification of questionnaires (which are very extensively adapted even when sharing a common language), English and French versions benefited from a co-ordination process similar to the one implemented for cognitive materials. A list of "tips" for verification of questionnaires, including

spelling, possibly recurring adaptation issues, and especially errata, errata candidates and “quasi-errata” (suggestions for improving the source) was maintained, built up, and used in each successive verification.

There was also an increased effort, versus previous PISA rounds, to harmonise the verification feedback for different language versions of questionnaires used in the same country (e.g. German, French and Italian for Switzerland, or the five language versions for Spain). Such versions are by necessity entrusted to different verifiers, but insofar as possible, cApStAn’s verification reviewers made a point of reviewing and delivering such versions together, striving to harmonise e.g. verification interventions on adaptation issues common to the different language versions.

Main Survey verification of questionnaires

Similarly to the procedure adopted as of 2009, to significantly save time during the reviewing process, ACER provided a customised QAS for each language version in which all the Field Trial national adaptations were imported for items that were left unchanged from the Field Trial.

The Main Survey QAS was designed to track the particular Main Survey translation/adaptation and verification process. Next to the Field Trial column documenting the English translation of the national version in Field Trial (which was locked), a column was added in which the National Centre was required to either confirm the unaltered Field Trial version or record the intended revised adaptation for Main Survey. If a source element had been altered by the Consortium, or if it was “dodgy”, i.e. it worked in an unexpected way during the Field Trial at national or international level, the National Centre was required to provide the Main Survey adaptation. In the QAS, source items removed from the Field Trial version were deleted without tracking, but for the items that were modified; “strikethrough” and coloured text were used to indicate the changes. Rows subject to verification were identified via colour-shading.

Once the negotiation of Main Survey adaptations was concluded, cApStAn could launch the verification. For the Main Survey (focused verification), the verifiers were instructed to: check that all the Consortium changes are correctly implemented, check that all national changes are appropriate and implemented as documented, verify dodgy items completely, re-check whether confirmed adaptations are correctly implemented, verify completely questions that were modified or added; and fully (re-)verify a selection of items.

Similarly to the process for paper-based cognitive materials, verifiers entered corrections in track changes mode in the Word® files and documented the verification in the QAS. The QAS was used like the TAS for paper-based units for a key corrections process (but using MyPISA for various stages).

As was the case for cognitive materials, verification revealed occurrences of undocumented Field Trial to Main Survey changes or Main Survey changes introduced in non-finalised (e.g. pre-verification) Field Trial instruments. Such occurrences were treated on a case per case basis. Verifiers re-implemented the still relevant Field Trial key corrections in the Main Survey version, and then performed the verification of target versus source as per Main Survey procedure.

Final (Optical) Check of test booklets, questionnaire booklets, computer-based test units, coding guides

As in all previous rounds, test booklets and questionnaire forms were checked, at both Field Trial and Main Survey, page-by-page for correct item allocation, layout, page numbering, item numbering, graphic elements, item codes and footers. This phase continues to prove essential in locating residual flaws, some of which occur only during assembly of materials.

The final check process for test booklets and questionnaire booklets in the PISA 2012 Main Survey was mostly unchanged from the Field Trial and from PISA 2009, though a more formalised Final Optical Check (FOC) review process was put in place. A FOC reviewer carefully reviews the draft FOC report (systematic check of all comments listed by the FOC verifier) plus transfers any issues concerning key corrections from the various adaptation spreadsheets to the FOC report, using the category “Residual issue at content level” (since PISA 2009, this is the process used to enforce key corrections). The FOC review process was formalised as a distinctive step in the PISA 2012 Main Survey, along the lines of what was done for paper-based clusters, computer-based units, questionnaires, and coding guides: a reviewer stands behind each verifier.

In practice, the FOC review leads to a “lighter” FOC report. FOC verifiers are trained to spot and report every possible layout deviation, and they do not necessarily have knowledge on specific languages nor are they acquainted with

all verification steps, in particular with the convergence check of link items. Thus a major aspect of FOC review is the "weeding out" or "qualification" of comments. In particular, "qualification" is often required when it comes to comments concerning link items, for which the FOC may reveal deviations versus the source version – rather than versus the previous national version. When minor issues or deviations had previously been approved (or had previously passed unchallenged) in link items, the FOC reviewer suitably modifies the comment.

Computer-based units were also subject to a final check process focusing on the correct implementation of key corrections. After verification and post-verification review by the country (including possible negotiation of key corrections with the Consortium referee), a final check reviewer consults the differences report and launches previews of the latest target versions uploaded by the country and of the English (or French) source versions as needed. If all key corrections are found to be correctly implemented, the final check reviewer documents this in the differences report and finalises the unit on the TMS, indicating it is ready for national construction. If a "key correction" is found not implemented (or only partially implemented, or defectively implemented), the final check reviewer describes the issue in the differences report and "un-finalises" the unit on the TMS, making it accessible again to the National Centre for re-editing.

For CBA units (as for Questionnaires but differently from PBA clusters, coding guides, booklets), the final check was part of a MyPISA task.

For the Main Survey (only), after verification and post-verification processing by the National Centres (including possible negotiation of key corrections with the Consortium referee), it was required to submit coding guides in PDF® format for a final check.

The final check of coding guides consisted in: a page-by-page check for completeness; a systematic check of correct implementation of key corrections from the verification of the general introductions and the verification and/or convergence check of coding sections; a random check of key corrections from the verification of stimuli + items, to determine whether final versions of stimuli and items were correctly used to assemble the coding guides (this part of the process was sometimes carried out at verification phase); a random check of late Consortium Field Trial changes in the coding section of new mathematics units, to determine whether the final Field Trial mathematics coding guide was correctly used as basis for updating to Main Survey – not forgetting that for most countries, the correct implementation of such changes was not checked at Field Trial when cognitive materials were verified in unit format, before the release of the final post-coder-training-meeting Field Trial coding guides (this part of the process was also sometimes carried out at verification phase).

The final check of coding guides did not include a layout check as such. Different from test booklets, countries are not required to echo the source version pagination and may even change the order of presentation of the material (e.g. some countries prefer cluster order rather than the unit order used in the source). Nonetheless, the completeness check often reveals minor layout or formatting defects (e.g. incorrect formatting of credit labels, descriptors, or sample student responses), which are then reported. In the case of one national version with changed order of presentation (Belgium-Dutch), it revealed an entire missing unit in the paper-based mathematics coding guide.

A final check report in Excel® format was sent to each National Centre, detailing any recommended corrections; these were infrequent and mostly minor (see above). Differently from the final check of test booklets or questionnaires, the procedure for coding guides was kept lighter in that National Centres were not asked to return the report with confirmations of their follow-up (selection for each issue of "Corrected" or "Not corrected" plus explanation in the latter case). In process terms it may be questioned whether this differential treatment is justified.

Similar to analyses of the test adaptation spreadsheet, cApStAn conducted quantitative analyses of FOC reports both at Field Trial and Main Survey phases, which gave good estimates of the number and types of residual errors found in assessment booklets.

Verification of operational manuals

The verification process for Manuals in PISA 2012 is basically unchanged from PISA 2009: ACER vets the national adaptations, following which cApStAn verifies a list of key components called "specified parts". One difference is that in 2012 it was decided that cApStAn would perform its verification only in the Main Survey – whereas in 2009 verification by cApStAn was performed in the Field Trial and then repeated in the Main Survey for "problematic countries", i.e. those for which significant manuals-related issues were identified at FT.

In practice the bulk of manuals verification work is carried out or anyway extensively prepared by cApStAn staff, with national verifiers consulted only as needed. Extensive explanations have been provided in previous verification reports on the reasons for involving verifiers the least possible in manuals verification.

Quantitative analyses of verification outcomes

In PISA 2000 and PISA 2003, verification reports contained qualitative information about the national versions and illustrative examples of typical errors encountered by the verifiers. As of the PISA 2006 Main Survey, the instruments used to document the verification were designed to generate statistics, and some quantitative data is available. The verification statistics by item and by unit yielded information on translation and adaptation difficulties encountered for specific items in specific languages or groups of languages. This type of information, when gathered during the Field Trial, could be instrumental in revising items for the Main Survey but would also give valuable information on how to avoid such problems in further survey administrations. The verification report includes all data and country names and is a confidential document reviewed by the technical advisory group. Each country received its own report and data.

This information also makes it possible to detect whether there are items that elicited many verifier interventions in almost all language groups. When this occurs, item developers would be prompted to re-examine the item's reliability or relevance. Similarly, observing the number of adaptations that the countries proposed for some items may give the item developers additional insight into how difficult it is for some countries to make the item suitable for their students. While such adaptations may be discussed with the Consortium, it remains likely that extensively adapted items will eventually differ from the source version (e.g. in terms of reading difficulty).

The verification reports for the PISA 2012 Field Trial and PISA 2012 Main Survey include sections with quantitative analyses conducted on verification and assessment booklet FOC outcomes. They also contain pointers and directions for further work that could be carried out in this direction. NPMs have shown a keen interest in this type of analysis.

SUMMARY OF ITEMS DELETED AT THE NATIONAL LEVEL, DUE TO TRANSLATION, PRINTING OR LAYOUT ERRORS

In all cases when serious flaws were identified in the functioning of specific items as administered in individual countries, the NPMs were asked to review their translation of the item and to provide the Consortium with possible explanations of the malfunction.

Across all administrations of the mathematics, reading, science, problem solving and financial literacy items used for PISA 2012, approximately 47 instances were identified for which data were excluded from the analysis for reasons related to a translation issue (38 cases), printing issue (3 cases) or layout issue (6 cases).

Some 24 countries had one or more items affected by such errors, and an additional three instances affected several countries using a common language (two items affecting several of the German-speaking countries, and one item affecting some of the countries using Chinese).

Notes

1. International Association for the Evaluation of Educational Achievement.

2. Available at *www.oecd.org/pisa*.

3. Available at *www.oecd.org/pisa*.

4. Available at *www.oecd.org/pisa*.

5. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

6. The MyPISA website contains the source versions of instruments, manuals and other documents and information relating to National Centres.

7. See note 5.

8. See note 5.

References

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

Thorndike, R.L. (1973), *Reading Comprehension in Fifteen Countries*, Wiley, New York, and Almqvist & Wiksell, Stockholm.

6

Field Operations

OVERVIEW OF ROLES AND RESPONSIBILITIES

PISA was co-ordinated in each country by a National Project Manager (NPM) who implemented the procedures specified by the international contractor responsible for PISA implementation. Each NPM typically had several assistants, working from a base location that is referred to throughout this report as a National Centre. For the school level operations the NPM co-ordinated activities with school level staff, referred to in PISA as School Co-ordinators. Trained Test Administrators administered the PISA assessment in schools.

National Project Managers

NPMs were responsible for implementing the project within their own country. They:

- attended NPM meetings and received training in all aspects of PISA operational procedures;
- negotiated nationally specific aspects of the implementation of PISA with the international contractor, such as national and international options, oversampling for regional comparisons, additional analyses and reporting, e.g. by language group;
- established procedures for the security and prospecting confidentiality of materials during all phases of the implementation;
- prepared a series of sampling forms documenting sampling related aspects of the national educational structure;
- prepared the school sampling frame and submitted this to the international contractor for the selection of the school sample;
- organised for the preparation of national versions of the test instruments, questionnaires, manuals and coding guides;
- identified School Co-ordinators from each of the sampled schools (nominated by the school principal or a volunteer from the school staff) and worked with them on school preparation activities;
- selected the student sample from a list of eligible students provided by the School Co-ordinators;
- recruited and trained Test Administrators according to the Technical Standards for PISA 2012, Standards 6.1, 6.2, and 6.3 to administer the tests within schools (see Annex F);
- nominated suitable persons to work on behalf of the international contractor as external quality monitors to observe the test administration in a selection of schools;
- recruited and trained coders to code the open-ended items;
- arranged for the data entry of the test and questionnaire responses, and submitted the national database of responses to the international contractor; and
- submitted a written review of PISA implementation activities following the assessment.

A *National Project Manager's Manual* provided detailed information about the duties and responsibilities of the NPM. Supplementary manuals, with detailed information about particular aspects of the project, were also provided. These included:

- a *School Sampling Preparation Manual,* which provided instructions to the NPM for documenting school sampling related issues such as the definition of the target population, school level exclusions, the proportion of small schools in the sample and so on. Instructions for the preparation of the sampling frame, i.e. the list of all schools containing PISA eligible students, were detailed in this manual; and
- a *Data Management Manual*, which described all aspects of the use of *KeyQuest*, the data entry software prepared by ACER for the data entry of responses from the tracking instruments, test booklets and questionnaires.

School Co-ordinators

School Co-ordinators co-ordinated school-related activities with the National Centre and the Test Administrators.

The School Co-ordinator :

- established the testing date and time in consultation with the NPM;
- prepared the student listing form with the names of all eligible students in the school and sent it to the NPM so that the NPM could select the student sample;

- received the list of sampled students on the student tracking form from the NPM and updated it if necessary, including identifying students with disabilities or limited test language proficiency who could not take the test according to criteria established by the international contractor;
- received, distributed and collected the School Questionnaire;
- received and distributed the Parent Questionnaire in the countries that implemented this international option (the Test Administrator distributes the Parent Questionnaire to students on the assessment day or 1-2 weeks before the assessment to deliver it to the parents to complete);
- informed school staff, students and parents of the nature of the test and the test date by sending a letter or organising a meeting, and secured parental permission if required by the school or education system;
- informed the NPM and Test Administrator of any test date or time changes; and
- assisted the Test Administrator with room arrangements for the test day.

On the test day, the School Co-ordinator was expected to ensure that the sampled students attended the test session(s). If necessary, the School Co-ordinator also made arrangements for a follow-up session and ensured that absent students attended the follow-up session.

A School Co-ordinator*'s Manual* was prepared by the international contractor, that described in detail the activities and responsibilities of the School Co-ordinator.

Test Administrators

The Test Administrators were primarily responsible for administering the PISA test fairly, impartially and uniformly, in accordance with international standards and PISA procedures. To maintain fairness, a Test Administrator could not be the reading, mathematics or science teacher of the students being assessed and it was preferred that they not be a staff member at any participating school (see Standards 6.1, 6.2, and 6.3 in Annex F). Prior to the test date, Test Administrators were trained by National Centres. Training included a thorough review of the *Test Administrator's Manual,* prepared by the international contractor, and the script to be followed during the administration of the test and questionnaire. Additional responsibilities included:

- ensuring receipt of the testing materials from the NPM and maintaining their security;
- co-operating with the School Co-ordinator ;
- contacting the School Co-ordinator one to two weeks prior to the test to confirm plans;
- completing final arrangements on the test day;
- conducting a follow-up session, if needed, in consultation with the School Co-ordinator ;
- reviewing and updating the student tracking form (a form designed to record sampled students with their background data);
- completing the session attendance form (a form designed to record sampled students attendance and instruments allocation), and the session report form (a form designed to summarise session times, any disturbance to the session, etc.);
- ensuring that the number of tests and questionnaires collected from students tallied with the number sent to the school;
- obtaining the School Questionnaire from the School Co-ordinator ; and
- sending the School Questionnaire, the Student Questionnaires and all test materials (both completed and not completed) to the NPM after the testing was carried out.

School associates

In some countries, one person undertook the roles of both School Co-ordinator and Test Administrator. In these cases, the person was referred to as the school associate and the same Standards 6.1, 6.2, and 6.3 apply as for the Test Administrator. A *School Associate's Manual* was prepared by the international contractor, combining the source material provided in the individual School Co-ordinator and Test Administrator manuals to describe in detail the activities and responsibilities of the school associate.

THE SELECTION OF THE SCHOOL SAMPLE

NPMs used the detailed instructions in the *School Sampling Preparation Manual* to document their school sampling plan and to prepare their school sampling frame.

The national target population was defined, school and student level exclusions were identified, and aspects such as the extent of small schools (a small school is defined as any school whose approximate enrolment falls below the target cluster size of 35 students, or fewer than target cluster size of 35 students plus 8 students if doing financial literacy) and the homogeneity of students within schools were considered in the preparation of the school sampling plan.

For all but a small number of countries, the sampling frame was submitted to the international contractor who selected the school sample. Having the international contractor select the school sample minimised the potential for errors in the sampling process, and ensured uniformity in the outputs for more efficient data processing later (student sampling, data analysis). It also relieved the burden of this task from National Centres. NPMs worked very closely with the international contractor throughout the process of preparing the sampling documentation, ensuring that all nationally specific considerations related to sampling were thoroughly documented and incorporated into the school sampling plan.

All countries were required to thoroughly document their school sampling plan. If there was any deviation noted the National Centre was required to explain in detail the sampling methods used, to ensure that they were consistent with those used by the international contractor. In this case, the standard procedure the international contractor used to check that the national school sampling had been implemented correctly was to draw a parallel sample using its international procedures and compare the two samples. Further details about sampling for the Main Survey are provided in Chapter 4.

PREPARATION OF TEST BOOKLETS, QUESTIONNAIRES AND MANUALS

As described in Chapter 2, 13 different test booklets had to be assembled with clusters of test items arranged according to the test booklet design specified by the international contractor. Test items were presented in units (stimulus material and items relating to the stimulus) and each cluster contained several units. Test units and questionnaire items were initially sent to NPMs several months before the testing dates, allowing adequate time for items to be translated. Units allocated to clusters and clusters allocated to booklets were provided a few weeks later, together with detailed instructions to NPMs about how to assemble their translated or adapted clusters into booklets.

For reference, source versions of all booklets were provided to NPMs in both English and French and were also available through a secure website. NPMs were encouraged to use the cover design provided by the OECD. In formatting translated or adapted test booklets, they had to follow as far as possible the layout of the source versions, including allocation of items to pages.

NPMs were required to submit their cognitive material in units, along with a form documenting any proposed national adaptations for verification by the international contractor. NPMs incorporated feedback from the verifier into their material and assembled the test booklets. These were submitted once more to the international contractor, who performed a Final Optical Check of the materials. This was a verification of the layout, instructions to the student, the rendering of graphic material, etc. Once feedback from the final optical check had been received and incorporated into the test booklets, the NPM was ready to send the materials to print.

The Student Questionnaire contained one or two modules, according to whether the Information Communication Technology familiarity questionnaire or the educational career questionnaire components were being added to the core component. Forty-two countries administered the information communication technology familiarity questionnaire and 23 countries administered the questionnaire on educational career.

The core component had to be presented first in the questionnaire booklet.

Eleven countries also administered the optional Parent Questionnaire. The online administration of the School Questionnaire was implemented by nineteen countries.

As with the test material, source versions of the questionnaire instruments in both French and English were provided to NPMs for translation into the languages of the test.

NPMs were permitted to add questions of national interest as national options to the questionnaires. Proposals and text for these were submitted to the international contractor for approval as part of the process of reviewing adaptations to

the questionnaires. It was required that the additional material should be placed at the end of the international modules. The student questionnaire was modified more often than the school questionnaire.

NPMs were required to submit a form documenting all proposed national adaptations to questionnaire items to the international contractor for approval. Following approval of adaptations, the material was verified by the international contractor. NPMs implemented feedback from verification in the assembly of their questionnaires, which were submitted once more in order to conduct a final optical check of the layout, etc. Following feedback from the Final Optical Check, NPMs made final changes to their questionnaires prior to printing.

The *School Co-ordinator's Manual* and *Test Administrator's Manual* (or the *School Associate Manual* for those countries that combined the roles of the School Co-ordinator and Test Administrator) were also required to be translated into the languages of instructions. Only English source versions of each manual were provided by the international contractor. NPMs were required to submit a form documenting all proposed national adaptations to the manuals to the international contractor for approval. Following approval of the adaptations, the manuals were prepared and submitted to the international contractor. A verification of key elements called 'specified parts' of the manuals – those related to the coding of the tracking instruments and the administration of the test – was conducted. NPMs implemented feedback from the verifier into their manuals prior to printing. A Final Optical Check was not required for the manuals.

In countries with multiple languages, the test instruments and manuals needed to be translated into each test language. For a small number of countries, where Test Administrators were bilingual in the test language and the national language, it was not required for the whole of the manuals to be translated into both languages. However in these cases it was a requirement that the test script, included within the Test Administrator manual was translated into the language of the test.

THE SELECTION OF THE STUDENT SAMPLE

Following the selection of the school sample by the international contractor, the list of sampled schools was returned to National Centres. NPMs then contacted these schools and requested a list of all PISA-eligible students from each school. This was provided on the *List of Students*, and was used by NPMs to select the student sample.

NPMs were required to select the student sample using *KeyQuest*, the PISA student sampling and data entry software prepared by the international contractor. *KeyQuest* generated the list of sampled students for each school, known as the *Student Tracking Form* and the *Session Attendance Form* that served as the central administration documents for the study and linked students, test booklets and Student Questionnaires.

PACKAGING AND SHIPPING MATERIALS

Regardless of how materials were packaged and shipped, the following needed to be sent either to the Test Administrator or to the school:

- test booklets and Student Questionnaires for the number of students sampled;
- student tracking form;
- two copies of the Session attendance form for both paper-based and computer-based assessments;
- two copies of the Session report form for both paper-based and computer-based assessments;
- test delivery USB sticks for the computer-based assessment;
- computer-based assessment Student logon form;
- results for the school's computer system diagnostic report to determine the suitability of running the computer-based assessment from USB;
- materials reception form;
- materials return form;
- additional materials, e.g., rulers, pens and calculators, as per local circumstances; and
- additional School and Student Questionnaires and a bundle of extra test booklets.

Of the 13 separate test booklets, one was pre-allocated to each student by the *KeyQuest* software from a random starting point in each school. *KeyQuest* was then used to generate the school's session attendance form, which contained the number of the allocated booklet alongside each sampled student's name.

It was recommended that labels be printed, each with a student identification number and test booklet number allocated to that identification, as well as the student's name if this was an acceptable procedure within the country. Two or three copies of each student's label could be printed, and used to identify the test booklet, the questionnaire, and a packing envelope if used.

NPMs were allowed some flexibility in how the materials were packaged and distributed, depending on national circumstances. It was specified however that the test booklets for a school be packaged so that they remained secure, possibly by wrapping them in clear plastic and then heat-sealing the package, or by sealing each booklet in a labelled envelope. Three scenarios, summarised here, were described as illustrative of acceptable approaches to packaging and shipping the assessment materials:

- Country A: All assessment materials shipped directly to the schools; school staff (not teachers of the students in the assessment) to conduct the testing sessions; materials assigned to students before packaging; materials labelled and then sealed in envelopes also labelled with the students' names and identification numbers.
- Country B: Materials shipped directly to the schools; external Test Administrators employed by the National Centre to administer the tests; the order of the booklets in each bundle matches the order on the session attendance form; after the assessment has been completed, booklets are inserted into envelopes labelled with the students' names and identification numbers and sealed.
- Country C: Materials shipped to Test Administrators employed by the National Centre; bundles of 35 booklets sealed in plastic, so that the number of booklets can be checked without opening the packages; Test Administrators open the bundle immediately prior to the session and label the booklets with the students' names and ID numbers from the student tracking form.

TEST ADMINISTRATION

On the assessment day after arriving at the school, Test Administrators were required to check whether the copy of the student tracking form was identical with the School Co-ordinator 's copy and then to set up the room and materials for the assessment session following the steps below:

- allocate a desk and seat to each participating student;
- prepare one test booklet (according to the previous random assignment specified on the session attendance form) and one Student Questionnaire for each student, labelled with the student's name and identification number;
- write the testing date on a board visible to all students;
- ask the students to write it on their test booklet covers in the required format DD/MM/YYYY at the beginning of the session; and
- set aside the materials for students who had any non-participant codes recorded on the student tracking form or did not attend the administration session from the very beginning.

To obtain comparable and reliable data Test Administrators were asked to strictly follow the anticipated timing of the paper-based assessment, especially the administration of the cognitive test booklet material (2 x 1 hour exactly) shown in Figure 6.1 below.

■ Figure 6.1 ■

Timing of test administration session

Activity	Time
Distributing the materials and reading the General Directions	10-15 minutes (approximately)
The test booklet	1 hour (exactly)
Short break	**<No more than 5 minutes>**
The test booklet	1 hour (exactly)
Break	**<15 minutes>**
The Student Questionnaire	<**35 minutes**> (approximately)
Collecting the materials and ending the session	3-5 minutes (approximately)
Total	**<3 hours and 15 minutes>** (approximately)

NPMs were allowed to adapt the length of the short break between the two 1-hour test booklet sessions referring to the national practice and policy. In a few cases countries preferred not to have any break during the Test Administration.

All suggested adaptations to the timing had to be negotiated and approved by ACER within the manuals' adaptation negotiation procedures. Adaptation to the timing of the Student Questionnaire session was more flexible in order to maximise the contextual data obtained from students.

Test Administrators were also responsible for monitoring the assessment session. They had to pay special attention to the student attendance. Students had to attend the beginning of the first part of the administration session to participate in the assessment and could not be admitted to the session once the booklet or questionnaire directions began. They were not allowed to leave the session unless it was absolutely necessary. In case a student could not complete the session for any reason the Test Administrator had to collect the student's test material. If the student was absent for more than 10 minutes from the test booklets session, the Test Administrator had to record this student as "partially present" on the session attendance form.

The script for spoken instructions had to be read to the students word-for-word to maintain standardised assessment procedures across all participating countries. Test Administrators were not allowed to provide any help with the cognitive items in the test booklet. However, they could answer questions about items in the student questionnaire. Observers were limited to necessary staff members and the PISA Quality Monitors if in attendance at the session.

At the end of the scheduled time when the paper-based administration was completed Test Administrators had to collect the assessment materials from all students as well as the completed School Questionnaire from the School Co-ordinator in case its paper-based form was administered in the school. The assessment material from each administration session had to be bundled together with the corresponding session attendance form, session report form, unused test booklets and student questionnaires and their spare copies prepared for shipment to the National Centre.

RECEIPT OF MATERIALS AT THE NATIONAL CENTRE AFTER TESTING

It was recommended that the National Centre establish a database of schools before testing began to record the shipment of materials to and from schools, tallies of materials sent and returned, and to monitor the progress of the materials throughout the various steps in processing booklets after the testing.

It was recommended that upon receipt of materials back from schools, the counts of completed and unused booklets also be checked against the participation status information recorded on the student tracking form by the Test Administrator.

CODING OF THE TESTS AND QUESTIONNAIRES

This section describes PISA's coding procedures, including multiple coding, and makes brief reference to pre-coding of responses to a few items in the Student Questionnaire. Overall, 27% of the cognitive items across the reading, mathematics, and science domains required manual coding by trained coders.

This was a complex operation, as booklets had to be randomly assigned to coders and, for the minimum recommended sample size per country of 5 200 students, more than 26 000 responses had to be evaluated. An average of five items from each of the thirteen booklets required evaluation.

It is crucial for comparability of results in a study such as PISA that students' responses are scored uniformly from coder to coder and from country to country. Comprehensive criteria for coding, including many examples of acceptable and unacceptable responses, were prepared by the international contractor and provided to NPMs in coding guides for each of the three core domains; mathematics, reading, science, and also the financial literacy as an optional domain.

Preparing for coding

In setting up the coding of students' responses to open-ended items, NPMs had to carry out or oversee several steps:

- adapt or translate the coding guides as needed and submit these to the international contractor for verification;
- recruit and train coders;
- locate suitable local examples of responses to use in training and practice;
- organise booklets as they were returned from schools;
- select booklets for multiple coding;

- do the single coding of booklets according to the international design (see Figures 6.2, 6.3, 6.4, 6.5, 6.6, 6.7 and 6.8); and
- do the multiple coding of a selected sub-sample of booklets for the reliability study according to the international design (see Figures 6.2, 6.9, 6.10 and 6.11) once the single coding was completed.

Detailed instructions for each step were provided in the *Procedures for Coding Paper-based Constructed-Response Items MS12 (Cdg_Procedures_PB_MS12_2.docx)*. Key aspects of the process are included here.[1]

International coder training

Representatives from each National Centre were required to attend two international coder training sessions – one immediately prior to the Field Trial and one immediately prior to the Main Survey. At the training sessions, international contractor staff familiarised National Centre staff with the coding guides and their interpretation.

Staffing

NPMs were responsible for recruiting appropriately qualified people to carry out the single and multiple coding of the test booklets. In some countries, pools of experienced coders from other projects could be called on. It was not necessary for coders to have high-level academic qualifications, but they needed to have a good understanding of either mid-secondary level mathematics and science or the language of the test, and to be familiar with ways in which secondary-level students express themselves. Teachers on leave, recently retired teachers and senior teacher trainees were all considered to be potentially suitable coders. An important factor in recruiting coders was that they could commit their time to the project for the duration of the coding, which was expected to take up to one month.

The international contractor provided a coder recruitment kit to assist NPMs in screening applicants. These materials were similar in nature to the coding guides, but were much briefer. They were designed so that applicants who were considered to be potentially suitable could be given a brief training session, after which they coded some student responses. Guidelines for assessing the results of this exercise were supplied. The materials also provided applicants with the opportunity to assess their own suitability for the task. The number of coders required was governed by the design for multiple coding (described in a later section). For the Main Survey, it was recommended to have 12 coders to code mathematics/financial literacy, 8 coders to code science, and additional 8 coders to code reading. The coding design was prepared for both standard and easier set of booklets, as detailed in a document *Paper-based coding design MS12* (*Cdg_design_PB_MS12_2*.xlsx).[2] These numbers of coders were considered to be adequate for countries testing between 4 500 (the minimum number required) and 6 000 students to meet the timeline of submitting their data within three months of testing.

For larger numbers of students or in cases where coders would code across different combinations of domains, NPMs could prepare their own design and submit it to the international contractor for approval. A minimum of four coders were required in each domain to satisfy the requirements of the multiple coding design. Given that several weeks were required to complete the coding, it was recommended that at least two back-up coders of mathematics/financial literacy, and one back-up science and reading coder be trained and included in at least some of the coding sessions.

The coding process was complex enough to require a full-time overall supervisor of activities who was familiar with the logistical aspects of the coding design, the procedures for checking coder reliability, the coding schedules and the content of the tests and coding guides.

NPMs were also required to designate persons with subject-matter expertise, familiarity with the PISA tests and, if possible, experience in coding student responses to open-ended items to act as table leaders during the coding. Table leaders were expected to participate in the actual coding and spend extra time monitoring consistency. Good table leaders were essential to the quality of the coding, as their main role was to monitor coders' consistency in applying the coding criteria. They also assisted with the flow of booklets, and fielded and resolved queries about the coding guide and about particular student responses in relation to the guide, consulting the supervisor as necessary when queries could not be resolved. The supervisor was then responsible for checking such queries with the international contractor.

People were also needed to unpack, check and assemble booklets into labelled bundles so that coders could respect the specified design for randomly allocating sets of booklets to coders.

Consortium coding query service

A coding query service was provided by the international contractor in case questions arose about particular items that could not be resolved at the National Centre. Responses to coding queries were placed on the website, accessible to the NPMs from all participating countries.

Confidentiality forms

Before seeing or receiving any copies of PISA test materials, prospective coders were required to sign a confidentiality form, obligating them not to disclose the content of the PISA tests beyond the groups of coders and trainers with whom they would be working.

National training

Anyone who coded the PISA Main Survey test booklets had to participate in specific training sessions, regardless of whether they had had related experience or had been involved in the PISA Field Trial coding. To assist NPMs in carrying out the training, the international contractor prepared training materials in addition to the detailed coding guides. Training within a country could be carried out by the NPM or by one or more knowledgeable persons appointed by the NPM. Subject matter knowledge was important for the trainer as was an understanding of the procedures, which usually meant that more than one person was involved in leading the training.

The recommended allocation of booklets to coders assumed coding by cluster. This involved completing the coding of each item separately within a cluster within all of the booklets allocated to the coder before moving to the next item, and completing one cluster before moving to the next.

Coders were trained by cluster for the seven mathematics clusters, two financial literacy clusters (in case they implemented the financial literacy option), the three science clusters and the three clusters of reading. During a training session, the trainer reviewed the coding guide for a cluster of units with the coders, and then had the coders assign codes to some sample items for which the appropriate codes had been supplied by the international contractor. The trainer reviewed the results with the group, allowing time for discussion, querying and clarification of reasons for the pre-assigned codes. Trainees then proceeded to code independently some local examples that had been carefully selected by the supervisor of coding in conjunction with National Centre staff. It was recommended that prospective coders be informed at the beginning of training that they would be expected to apply the coding guides with a high level of consistency, and that reliability checks would be made frequently by table leaders and the overall supervisor as part of the coding process.

Ideally, table leaders were trained before the larger groups of coders since they needed to be thoroughly familiar with both the test items and the coding guides. The coding supervisor explained these to the point where the table leaders could code and reach a consensus on the selected local examples to be used later with the larger group of trainees. They also participated in the training sessions with the rest of the coders, partly to strengthen their own knowledge of the coding guides and partly to assist the supervisor in discussions with the trainees of their pre-agreed codes to the sample items. Table leaders received additional training in the procedures for monitoring the consistency with which coders applied the criteria.

Length of coding sessions

Coding responses to open-ended items is mentally demanding, requiring a level of concentration that cannot be maintained for long periods of time. It was therefore recommended that coders work for no more than six hours per day on actual coding, and take two or three breaks for coffee and lunch. Table leaders needed to work longer on most days so that they had adequate time for their monitoring activities.

Logistics prior to coding

Sorting booklets

When booklets arrived back at the National Centre, they were first tallied and checked against the session participation codes on the session attendance form. Unused and used booklets were separated; used booklets were sorted by student identification number if they had not been sent back in that order and then were separated by booklet number; and school bundles were kept in school identification order, filling in sequence gaps as packages arrived. Session attendance forms were copied, and the copies filed in school identification order. If the school identification number order did not

correspond with the alphabetical order of school names, it was recommended that an index of school name against school identification be prepared and kept with the binders.

Because of the time frame within which countries had to have all their coding done and data submitted to the international contractor, it was usually impossible to wait for all materials to reach the National Centre before beginning to code. In order to manage the design for allocating booklets to coders, however, it was recommended to start coding only when at least half of the booklets had been returned.

Selection of booklets for multiple coding

Each country was required to set aside 100 of each booklet from a standard set of booklets (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13) or from an easier set of booklets (8, 9, 10, 11, 12, 13, 21, 22, 23, 24, 25, 26, 27) and 50 of each booklet from four financial literacy booklets (71, 72, 73, 74) if applicable for multiple coding for the main language. For the PISA Main Survey 2012, only items from the first cluster in each booklet were multiple coded. This meant that there were three clusters left over from these multiple coded booklets that needed to be single coded. Because of the complexity of the single coding operation the yellow and blue batches were introduced:

a) The batches of booklets selected for the single coding operation were called 'The Yellow Batches' and they were labelled by numbers 1 to 12.

b) The batches of booklets selected for the multiple coding operation were called 'The Blue Batches' and they were labelled by letters A, B, C, and D.

The main principle in setting aside the booklets for multiple coding was that the selection needed to ensure a wide spread of schools and students across the whole sample and to be random as far as possible. The simplest method for carrying out the selection was to use a ratio approach based on the expected total number of completed booklets.

In most countries, approximately 400 of each booklet were expected to be completed, so the selection of booklets to be set aside for multiple coding required that approximately one in four booklets was selected. Depending on the actual numbers of completed booklets received, the selection ratios needed to be adjusted so that the correct numbers of each booklet were selected from the full range of participating schools.

In a country where booklets were provided in more than one language, if the minority language represented 20% or more of the target population, the country had to include at least 50 of each booklet form of that verified language. Multiple coding was not required for minority languages representing less than 20% of the target population.

Booklets for single coding

Single coding was required for all clusters within booklets in the yellow batches (single coding stage 1) and for the second, third and fourth clusters within booklets in the blue batches selected for multiple coding (single coding at stage 2). Some items requiring coding did not need to be included in the multiple coding. These were closed constructed response items that required a coder to assign a right or wrong code, but did not require any coder judgement. The coders in the single-coding process at stage 2 coded these items in the booklets set aside for multiple coding, as well as the items requiring single coding from the remaining second, third, and fourth clusters. Other items such as multiple-choice response items required no coding and were directly data-entered.

How codes were shown

A string of small code numbers corresponding to the possible codes for the item as delineated in the relevant coding guide appeared in the upper right-hand side of each item in the test booklets. For booklets being processed by a single coder, the code assigned was indicated directly in the booklet by circling the appropriate code number alongside the item. Tailored coding record sheets were prepared for each booklet for the multiple coding and used by all but the last coder so that each coder undertaking multiple coding did not know which codes other coders had assigned.

For all reading clusters and the majority of the mathematics/financial literacy and science clusters, item codes were often just 0, 1, and 9, indicating incorrect, correct and missing, respectively. Provision was made for some of the open-ended items to be coded as partially correct, usually with '2' as fully correct and '1' as partially correct.

For the mathematics/financial literacy and science clusters, a two-digit coding scheme was adopted for the items requiring constructed responses (8% of mathematics items, 26% of financial literacy items, and 21% of science items). The first digit represented the degree of correctness code; the second indicated the content of the response or the type of solution method used by the student.

Coder identification numbers

Coder identification numbers were assigned according to a standard three-digit format specified by the international contractor. The first digit showed the combination of domains that the coder would be working across, and the second and third digits had to uniquely identify the coders within their set. For example, twelve coders coding across the domains of mathematics/financial literacy and science were given identification numbers 501 to 512. Eight coders who coded just science were given identification numbers 301 to 308. Coder identification numbers were used for two purposes: implementing the design for allocating booklets to coders and monitoring coder consistency in the multiple-coding exercises.

Coding Operation

The whole coding operation had four stages (see Figure 6.2).

■ Figure 6.2 ■

PISA 2012 Main Survey coding design

PISA 2012 MAIN SURVEY CODING DESIGN			
Single Coding (SC)		**Multiple Coding (MC)**	
Stage 1	**Stage 2**	**Stage 3**	**Stage 4**
Mathematics/financial literacy, science and reading clusters in yellow batches 1-12	Mathematics/financial literacy, science and reading clusters in blue batches A, B, C, D	Mathematics/financial literacy, science and reading clusters in blue batches A, B, C, D	Mathematics/financial literacy, science and reading clusters in blue batches A, B, C, D
Booklets selected for single coding	Booklets selected for multiple coding	Booklets selected for multiple coding	Booklets selected for multiple coding
Clusters 1, 2, 3, 4	Clusters 2, 3, 4	Cluster 1	Cluster 1
		Groups of 4 coders	Groups of 4 coders
		First three rounds of coding into multiple coding record sheets	Fourth round of coding directly into test booklets

The single coding consisted of two stages. In stage 1, coders worked only with the yellow batches 1 to 12 and they coded all mathematics/financial literacy, science and reading clusters from booklets selected only for single coding. In stage 2 coders worked only with the blue batches A, B, C, and D. They single coded all second, third, and fourth mathematics/ financial literacy, science and reading clusters from booklets selected only for multiple coding.

The multiple coding also consisted of two stages. In stage 3 coders worked only with the blue batches A, B, C, and D and they coded first mathematics/financial literacy, science and reading clusters from booklets selected only for multiple coding. Groups of four coders recorded the first three rounds of coding into multiple coding record sheets. In stage 4 coders worked only with the blue batches and again they coded first mathematics/financial literacy, science and reading clusters from booklets selected only for multiple coding. Groups of four coders recorded the fourth round of coding directly into the test booklets.

Single coding design

The design was organised so that all appearances of each cluster type involved in the single coding were coded together. This arrangement entailed coders working with several booklet types at the same time, and at times required space for partly coded booklets to be stored while other booklets were being worked on. However organising the coding this way had the substantial benefits of:

- obtaining more accurate and consistent coding (because training and coding are more closely linked); and
- minimising effects of coder leniency or harshness (more than one coder codes each booklet; coders code across the range of schools sampled).

Coding operation could be conducted in two waves. The first wave begins when, say, 60 % of the booklets are returned to the centre. After receiving all remaining 40 % booklets from schools the second wave begins.

Step 1 in Figure 6.3, for example, represents the training and coding sequence. Coding of all items in the cluster identified in one row should be completed before proceeding to training of the cluster identified in the following row. Each cluster from booklets 1-13 occurs in four booklets, and so several booklets are sometimes required for a coding step (i.e. a row of the Figure). Four booklets are included in the coding of PM1 cluster in this step. At stage 1 the blue batches are not used.

Once wave 1 is completed, and the remaining 40 % booklets are back, wave 2 of the single coding operation begins. At stage 1 each of the yellow batches is specifically allocated to a particular coder. At stage 2 the blue batches are needed. The familiar coding steps shown at each row of the Figure involve the single coding of clusters from the blue batches.

If wave 1 begins when 60% of the booklets have been returned to the centre, with a typical sample size of around 5200 students, there will be 3120 booklets coded during wave 1, so there will be about $3125/13 = 240$ of each booklet type. 60 of each booklet type will have been selected for the multiple coding (i.e. 60% of the 100 of each booklet type required). For the moment these are just set aside. (At the start of wave 2, when all 100 of each booklet type are available, these booklets are allocated into the blue batches.) The remaining 180 booklets will be allocated to the yellow batches. There are 12 single coding batches, so there should be around 15 books in each batch. Each coder is allocated 4 of these batches. For example coder 101 is allocated batch 1 of booklets 5, 10, 13 and 6. So each coder will have around 60 booklets.

Once the wave 1 single coding has been completed, i.e. all of the clusters from 180 x 13 booklet types in the yellow batches have been single coded, wave 2 begins.

In wave 2, there should be around 2080 booklets, around 160 of each booklet. 40 more of each booklet will be selected out for multiple coding to make up the 100 booklets required. Each of the 100 books for each booklet type selected for multiple coding are allocated into 4 blue batches of 25 books each. The remaining 120 booklets from wave 2 are allocated into 12 yellow batches (average of 10 per batch).

At stage 1 of wave 2 the coders get 4 yellow batches, so around 40 booklets to code. At stage 2, each blue batch has around 25 booklets, so the coding for this stage should be a little quicker than for the first stage.

Single coding of mathematics/financial literacy

In order to code by cluster, each coder needed to handle four of the thirteen booklet types at a time. For example, mathematics cluster 1 (PM1) occurred in booklets 5, 10, 13, and 6. Each of these occurrences had to be coded before another cluster was started. Moreover, since coding was done item by item, the item was coded across these different booklet types before the next item was coded.

A design to ensure the random allocation of booklets to coders was prepared based on the recommended number of 12 coders and the minimum sample size of 4 500 students from 150 schools. With 150 schools and 12 coders, each coder had to code a cluster within a booklet from twelve or thirteen schools ($150 / 12 \approx 13$). Figure 6.3 shows how booklets needed to be assigned to coders for the single coding. Further explanation of the information in this figure is presented below.

According to this design, cluster PM1 in yellow batch 1 (subset of schools 1 to 12) was to be coded by coder 101, cluster PM1 in yellow batch 2 (subset of schools 13 to 24) was to be coded by coder 102, and so on. For cluster PM2, coder 101 was to code all from yellow batch 2 (subset of schools 25 to 36) and coder 102 was to code all from batch 3 (subset of schools 37 to 48), and so on.

If booklets from all participating schools were available before the coding began, implementing this design involves the following steps at stage 1 (Figure 6.3). It is assumed here that training is conducted separately for each cluster prior to the start of its coding:

- The coders are trained in the coding of the items to be coded from cluster PM1.
- Coders then work through the locally prepared practice exercises. The coding of these items is monitored by the trainers and table leaders as described earlier.
- PM1 appears in booklets 5, 10, 13 and 6, so coders will be working with these four booklets at this step.
- Coder 101 takes batch 1 of booklets 5, 10, 13 and 6; Coder 102 takes batch 2 of these booklets, and so on through to Coder 112 who takes batch 12.

■ Figure 6.3 ■

Design for the single coding of mathematics/financial literacy, stage 1

SINGLE CODING STAGE 1

12 Mathematics/Financial Literacy coders

MATHEMATICS/FINANCIAL LITERACY CLUSTERS

				Yellow batches											
				1	2	3	4	5	6	7	8	9	10	11	12
	Step	Cluster	Booklets selected for single coding	Coder											
ALL COUNTRIES	1	PM1	5, 10, 13, 6	101	102	103	104	105	106	107	108	109	110	111	112
	2	PM2	6, 7, 11, 12	112	101	102	103	104	105	106	107	108	109	110	111
	3	PM3	3, 7, 9, 10	111	112	101	102	103	104	105	106	107	108	109	110
	4	PM4	4, 8, 10, 11	110	111	112	101	102	103	104	105	106	107	108	109
	5	PM5	1, 5, 9, 11	109	110	111	112	101	102	103	104	105	106	107	108
	6	PM6A	1, 3, 4, 6	108	109	110	111	112	101	102	103	104	105	106	107
	7	PM7A	2, 4, 5, 7	107	108	109	110	111	112	101	102	103	104	105	106
FL COUNTRIES	8	PM5	71, 72, 73, 74	106	107	108	109	110	111	112	101	102	103	104	105
	9	PF1	71, 72, 73, 74	105	106	107	108	109	110	111	112	101	102	103	104
	10	PF2	71, 72, 73, 74	104	105	106	107	108	109	110	111	112	101	102	103
FLUH COUNTRIES	11	PFUH	FLUH (70)	103	104	105	106	107	108	109	110	111	112	101	102
	12	PMUH	FLUH (70)	102	103	104	105	106	107	108	109	110	111	112	101
UH COUNTRIES	13	PMUH	UH (20)	101	102	103	104	105	106	107	108	109	110	111	112

- Coders then code all of the first PM1 item requiring coding in the booklets that they have.
- Note that PM1 appears in all four booklets, but in different locations within these four booklet types. So the question numbers for the same PM1 items will be different in these two booklet types. The same will be true for all clusters.
- Next, the second PM1 item is coded in each of the booklets held by the coder, followed by the third PM1 item, and so on until all of the PM1 items have been coded.
- Following the completion of this step (i.e. the first row), one PM1 cluster within the booklet 5's, 10's, 13's and 6's will have been coded.
- Training and then practice with local examples is then conducted in relation to cluster PM2.
- For the second step, booklets 6, 7, 11 and 12 are required. The booklets 5's, 10's and 13's used in the first step are not required for this step and so can be returned to the administration area.
- The batch 1 of booklet 6 that coder 101 used in the first step is now passed to coder 112. This coder is also provided with batch 1 of booklets 7, 11 and 12. Similarly, coder 101 receives the batch 2 of booklets 6 and also batch 2 of booklets 7, 11 and 12, coder 102 receives the batch 3 and so on, as shown in the second row.
- The items requiring coding from these clusters are coded item by item as described above, until all items have been coded.
- Training is now conducted for clusters PM3. Following training and practice using local examples, coder 111 takes batch 1 of booklets 3, 7, 9 and 10; coder 112 takes batch 2 of booklets 3, 7, 9 and 10, and so on according to the third row, and codes the items in the manner described above.
- The booklet batches should be kept intact with their batch header sheets throughout this operation. For some of the booklet types, the same batches will also be used during the multiple coding.

As a result of this procedure, the 12 mathematics/financial literacy coders will each process some booklets from seven of the twelve batches, and therefore will have coded across a wide range of schools. Each coder will have coded every mathematics/financial literacy cluster, and will therefore be well prepared for multiple coding.

At stage 2 the blue batches are needed. The familiar coding steps shown at each row of Figure 6.4 involve the single coding of clusters from the blue batches. There are only 12 blue batches to be coded for mathematics/financial literacy clusters at each step. For example, for cluster PM1, the batches needing coding are batches A-D of booklet 5; batches A-D of booklet 10; and batches A-D of booklet 13. Batches A-D of booklet 6 are not coded at this stage. They will be coded later, during the multiple coding operation.

■ Figure 6.4 ■

Design for the single coding of mathematics/financial literacy, stage 2

SINGLE CODING STAGE 2					
12 Mathematics/Financial Literacy coders					
MATHEMATICS/FINANCIAL LITERACY CLUSTERS					
				Blue batches	
				A B C D	
	Step	**Cluster**	**Booklets selected for multiple coding**	**Coder**	
ALL COUNTRIES	1	PM1	5, 10, 13	Any available Maths/FL coder	PM1 from the blue batches of booklet 6 is NOT coded until multiple coding
	2	PM2	6, 11, 12	Any available Maths/FL coder	PM2 from the blue batches of booklet 7 is NOT coded until multiple coding
	3	PM3	3, 7, 9	Any available Maths/FL coder	PM3 from the blue batches of booklet 10 is NOT coded until multiple coding
	4	PM4	4, 8, 10	Any available Maths/FL coder	PM4 from the blue batches of booklet 11 is NOT coded until multiple coding
	5	PM5	5, 9, 11	Any available Maths/FL coder	PM5 from the blue batches of booklet 1 is NOT coded until multiple coding
	6	PM6A	1, 3, 6	Any available Maths/FL coder	PM6A from the blue batches of booklet 4 is NOT coded until multiple coding
	7	PM7A	2, 4 ,7	Any available Maths/FL coder	PM7A from the blue batches of booklet 5 is NOT coded until multiple coding
FL COUNTRIES	8	PF1	72	Any available Maths/FL coder	PF1 from the blue batches of booklet 71 is NOT coded until multiple coding
	9	PF2	71	Any available Maths/FL coder	PF2 from the blue batches of booklet 72 is NOT coded until multiple coding

While the yellow batches are specifically assigned to coders, any available coder can be assigned the blue batches. Faster coders who finish stage 1 more quickly can be assigned one of the blue batches at stage 2 and do not need to wait for slower coders. If necessary, two (slower) coders could share a batch so that all 12 coders are occupied.

Single coding of science and reading

A similar design was prepared for the single coding of science and reading clusters. The same procedure applies at stage 1 described in the Figures 6.5, 6.6, 6.7 and 6.8. The recommended number of coders for each science (8) and reading (8) was one third less coders that recommended for coding mathematics items. First four coders (301-304) were allocated "two yellow batches' worth" of schools and last four coders (305-308) only "one yellow batch" at step 1. At step 2 in reverse, coders 301-304 were allocated one yellow batch, whereas coders 305-308 were allocated two yellow batches. Also, as there were just three different clusters of both reading and science, each of which appeared in nine booklet types, each coder coded all four appearances of a cluster. This ensured that a wider range of coders was used for each school subset. For the coding of cluster PS1, for example, coder 301 coded this cluster in booklets 12, 3, 5 and 8 from yellow batches 1 (i.e. schools 1-12) and 9 (i.e. schools 97-108) at step 1, then PS2 cluster in booklets 8, 1, 13 and 7 from yellow batch 5 (i.e. schools 49-60) at step 2, and finally PS3 cluster in booklets 10, 2, 1 and 12 from yellow batches 1 (i.e. schools 1-12) and 9 (i.e. schools 97-108). Coder 302 coded PS1 cluster from booklets 12, 3, 5 and 8 from yellow batches 2 (i.e. schools 13-24) and 10 (i.e. schools 109- 120) at step 1, then PS2 cluster in booklets 8, 1, 13 and 7 from yellow batch 6 (i.e. schools 61-72), and so on.

■ Figure 6.5 ■

Design for the single coding of science, stage 1

SINGLE CODING STAGE 1															
8 Science coders															
SCIENCE CLUSTERS															
				Yellow batches											
				1	2	3	4	5	6	7	8	9	10	11	12
	Step	**Cluster**	**Booklets selected for single coding**	**Coder**											
ALL COUNTRIES	1	PS1	12, 3, 5, 8	301	302	303	304	305	306	307	308	301	302	303	304
	2	PS2	8, 1, 13, 7	305	306	307	308	301	302	303	304	305	306	307	308
	3	PS3	10, 2, 1, 12	301	302	303	304	305	306	307	308	301	302	303	304
UH COUNTRIES	4	PSUH	UH (20)	305	306	307	308	301	302	303	304	305	306	307	308

Figure 6.6

Design for the single coding of science, stage 2

SINGLE CODING STAGE 2					
8 Science coders					
SCIENCE CLUSTERS					
				Blue batches	
				A B C D	
	Step	**Cluster**	**Booklets selected for multiple coding**	**Coder**	
ALL COUNTRIES	**1**	**PS1**	**3, 8, 5**	Any available Science coder	PS1 from the blue batches of booklet 12 is NOT coded until multiple coding
	2	**PS2**	**1, 7, 13**	Any available Science coder	PS2 from the blue batches of booklet 8 is NOT coded until multiple coding
	3	**PS3**	**1, 10, 12**	Any available Science coder	PS3 from the blue batches of booklet 2 is NOT coded until multiple coding

Figure 6.7

Design for the single coding of reading, stage 1

SINGLE CODING STAGE 1															
8 Reading coders															
READING CLUSTERS															
				Yellow batches											
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
	Step	**Cluster**	**Booklets selected for single coding**	**Coder**											
ALL COUNTRIES	**1**	**PR1**	**13, 12, 4, 9**	201	202	203	204	205	206	207	208	201	202	203	204
	2	**PR2**	**9, 2, 6, 8**	205	206	207	208	201	202	203	204	205	206	207	208
	3	**PR3**	**2, 11, 13, 3**	201	202	203	204	205	206	207	208	201	202	203	204
FL COUNTRIES	**4**	**PR2**	**71, 72, 73, 74**	205	206	207	208	201	202	203	204	205	206	207	208
UH COUNTRIES	**5**	**PRUH**	**UH (20)**	201	202	203	204	205	206	207	208	201	202	203	204

Figure 6.8

Design for the single coding of reading, stage 2

SINGLE CODING STAGE 2					
8 Reading coders					
READING CLUSTERS					
				Blue batches	
				A B C D	
	Step	**Cluster**	**Booklets selected for multiple coding**	**Coder**	
ALL COUNTRIES	**1**	**PR1**	**4, 9, 12**	Any available Reading coder	PR1 from the blue batches of booklet 13 is NOT coded until multiple coding
	2	**PR2**	**2, 8, 6**	Any available Reading coder	PR2 from the blue batches of booklet 9 is NOT coded until multiple coding
	3	**PR3**	**2, 11, 13**	Any available Reading coder	PR3 from the blue batches of booklet 3 is NOT coded until multiple coding

Countries implementing the optional UH booklet

Countries using the shorter, special purpose UH booklet (the *une heure* booklet see Chapter 2 for more details on UH booklet) were advised to process this separately from the remaining booklets. Small numbers of students used this booklet, only a few items required coding, and they were not arranged in clusters. NPMs were cautioned that booklets needed to be allocated to several coders to ensure uniform application of the coding criteria for UH booklet, as for the main coding.

Multiple coding

For PISA 2012, all booklets types, test booklets 1-13 for the standard set and test booklets 8-13 and 21-27 for the easier set, were involved in the multiple coding exercise. The first of the four clusters from all booklets were each independently coded by four separate coders according to the recommended design. The other three clusters from these booklets were already coded as part of the single coding design at stage 2 discussed above.

Multiple coding was done at or towards the end of the coding period, after coders had familiarised themselves with and were experienced in using the coding guides. As noted earlier, the first three coders of the selected booklets circled codes on separate record sheets, tailored to booklet type and domain (mathematics/financial literacy, science or reading), using one page per student. The coding supervisor checked that coders correctly entered student identification numbers and their own identification number on the sheets, which was crucial to data quality. The UH booklet was not included in the multiple coding.

While coders would have been thoroughly familiar with the coding guides by the time of multiple coding, they may have most recently coded a different booklet from those allocated to them for multiple coding. For this reason, they needed to have time to re-read the relevant coding guide before beginning the coding. It was recommended that time be allocated for coders to refresh their familiarity with the guides and to look again at the additional practice material before proceeding with the multiple coding. As in the single coding, coding was to be done item by item. For manageability, items from the first clusters within a booklet type were coded before moving to another booklet type, rather than coding by cluster across several booklet types. It was considered that coders would be experienced enough in applying the coding criteria by this time that coding by booklet would be unlikely to detract from the quality of the data.

Multiple coding of mathematics/financial literacy

The specified multiple coding design for mathematics/financial literacy, shown in Figure 6.9 assumed 12 coders with identification numbers 101 to 112. The importance of following the design exactly as specified was stressed, as it provided for links between clusters and coders. Figure 6.9 shows 12 coders grouped into four groups of four, with Group 1 comprising the first four coders (101-104), Group 2 the next four (105-108), and Group 3 the remaining four coders (109-112). The four codings were to be carried out by rotating the booklets to the four coders assigned to each group.

■ Figure 6.9 ■

Design for the multiple coding of mathematics/financial literacy, stages 3 and 4

MULTIPLE CODING STAGES 3 and 4				
12 Mathematics/Financial literacy Coders				
MATHEMATICS/FINANCIAL LITERACY CLUSTERS				
				Blue batches
				A B C D
	Step	Clusters for multiple coding	Booklets selected for multiple coding	Coder IDs
ALL COUNTRIES	1	PM1	6	101, 102, 103, 104
		PM2	7	105, 106, 107, 108
		PM3	10	109, 110, 111, 112
	2	PM4	11	101, 102, 103, 104
		PM5	1	105, 106, 107, 108
		PM6A	4	109, 110, 111, 112
	3	PM7A	5	101, 102, 103, 104
FL COUNTRIES		PF1	71	105, 106, 107, 108
		PF2	72	109, 110, 111, 112

In this scenario, with all 12 coders working, booklets 6, 7 and 10 were to be coded at the same time in the first step. The 100 booklet 6's, for example, were to be divided into four bundles of 25 and rotated among coders 101, 102, 103 and 104, so that each coder eventually would have coded clusters PM1 from all of the 100 booklets. Similarly, coders 105, 106, 107 and 108 coded PM2 cluster in booklets 7, and coders 109, 110, 111 and 112 coded PM3 cluster in booklets 10 in the first step. As described earlier, the first three coders recorded their codes on the separate multiple coding record sheets, while the fourth coder recorded his or her codes in the booklets themselves. The fourth coder had to also record his or her coder ID on the front cover of the booklet. After booklets 6, 7 and 10 had been put through the multiple-coding process, Group 1 (coders 101, 102, 103, 104) continued with coding of PM4 cluster in booklets 11, Group 2 (coders 105, 106, 107, 108) with PM5 in booklets 1, and Group 3 (coders 109, 110, 111, 112) with PM6A in booklets 4. Allocating booklets to coders for multiple coding was quite complex and the coding supervisor had to monitor the flow of booklets throughout the process.

Multiple coding of science and reading

The multiple-coding design for science shown in Figure 6.10 assumed eight coders with identification numbers 301 to 308 and for reading shown in Figure 6.11 assumed also eight coders with identification numbers 201 to 208.

■ Figure 6.10 ■

Design for the multiple coding of science, stages 3 and 4

MULTIPLE CODING STAGES 3 and 4				
8 Science coders				
SCIENCE CLUSTERS				
				Blue batches
				A B C D
	Step	**Clusters for multiple coding**	**Booklets selected for multiple coding**	**Coder IDs**
ALL COUNTRIES	1	PS1	12	301, 302, 303, 304
		PS2	8	305, 306, 307, 308
	2	PS3	2	301, 302, 303, 304

■ Figure 6.11 ■

Design for the multiple coding of reading, stages 3 and 4

MULTIPLE CODING STAGES 3 and 4				
8 Reading coders				
READING CLUSTERS				
				Blue batches
				A B C D
	Step	**Clusters for multiple coding**	**Booklets selected for multiple coding**	**Coder IDs**
ALL COUNTRIES	1	PR1	13	201, 202, 203, 204
		PR2	9	205, 206, 207, 208
	2	PR3	3	201, 202, 203, 204

If different coders were used for science or reading, a different multiple-coding design was necessary. The NPM would negotiate a suitable proposal with the international contractor. The minimum allowable number of coders coding a domain was four; in this case each booklet had to be coded by each coder.

Managing the coding process

Booklet flow

To facilitate the flow of booklets, it was important to have ample table surfaces on which to place and arrange them by type and school subset. The bundles needed to be clearly labelled. For this purpose, it was recommended that each bundle of booklets be identified by a batch header for each booklet type (standard set of booklets 1 to 13, easier set of booklets 8 to 13 and 21 to 27), with spaces for the number of booklets and school identification numbers in the bundle to be written in. In addition, each header sheet was to be pre-printed with a list of the clusters in the booklet, with columns alongside which the date and time, coder's name and identification number, and table leader's initials could be entered as the bundle was coded and checked.

Separating the coding of mathematics/financial literacy, science and reading

While consideration of the possibility that coders from different domains would require the same booklets at the same time was factored into the design of the single coding scheme, there was still the potential for this clash to occur. To minimise the risk of different coders requiring the same booklets, so that an efficient flow of booklets through the coding process could be maintained, it was recommended that the coding of mathematics/financial literacy and the coding of science and reading be done at least partly at different times (for example, mathematics/financial literacy coding could start a week or two ahead).

Familiarising coders with the coding design

The relevant design for allocating booklets to coders was explained either during the coder training session or at the beginning of the first coding session (or both). The coding supervisor was responsible for ensuring that coders adhered to the design and used clerical assistants if needed. Coders could better understand the process if each was provided with a card indicating the bundles of booklets to be taken and in which order.

Consulting table leaders

During the initial training, practice and review, it was expected that coding issues would be discussed openly until coders understood the rationale for the coding criteria (or reached consensus where the coding guide was incomplete). Coders were not permitted to consult other coders or table leaders during the additional practice exercises (see next subsection) undertaken following the training to gauge whether all or some coders needed more training and practice.

Following the training, coders were advised to work quietly, referring queries to their table leader rather than to their neighbours. If a particular query arose often, the table leader was advised to discuss it with the rest of the group.

For the multiple coding, coders were required to work independently without consulting other coders.

Monitoring single coding

The steps described here represented the minimum level of monitoring activities required. Countries wishing to implement more extensive monitoring procedures during single coding were encouraged to do so.

The supervisor, assisted by table leaders, was advised to collect coders' practice papers after each cluster practice session and to tabulate the codes assigned. These were then to be compared with the pre-agreed codes: each matching code was considered a hit and each discrepant code was considered a miss. To reflect an adequate standard of reliability, the ratio of hits to the total of hits plus misses needed to be 0.85 or more. This reliability was to be assessed on the first digit of the two-digit codes where applicable (some mathematics/financial literacy and science items). A ratio of less than 0.85, especially if lower than 0.80, was to be taken as indicating that more practice was needed, and possibly more training.

Table leaders played a key role during each coding session and at the end of each day, by spot-checking a sample of booklets or items that had already been coded to identify problems for discussion with individual coders or with the wider group, as appropriate. All booklets that had not been set aside for multiple coding were candidates for this spot-checking. It was recommended that, if there were indications from the practice sessions that one or more particular coders might be consistently experiencing problems in using the coding guide, then more of those coders' booklets should be included in the checking. Table leaders were advised to review the results of the spot-checking with the coders at the beginning of the next day's coding. This was regarded primarily as a mentoring activity, but NPMs were advised to keep in contact with table leaders and the coding supervisor if there were individual coders who did not meet criteria of adequate reliability and would need to be removed from the pool.

Table leaders were to initial and date the header sheet of each batch of booklets for which they had carried out spot-checking. Some items/booklets from each batch and each coder had to be checked.

Cross-national bias analysis

Cross-national comparability in assigning codes was explored through an inter-country coder reliability study (see Chapter 13).

Questionnaire coding

The main coding required for the Student Questionnaire internationally was the mother's and father's occupation. Four-digit International Standard Classification of Occupations (ISCO-08) codes (International Labour Organisation, 2007) were assigned to these two variables. In several countries, this could be done in a number of ways. NPMs could use a national coding scheme with more than 100 occupational title categories, provided that this national classification could be recoded to ISCO. In majority of cases (84%) ISCO-08 was the countries' choice of the occupational classification.

Given the update from ISCO-88 to ISCO-08 in PISA 2012, in addition to specific training sessions at NPM and coder meetings, NPMs were sent information on the structure of the new ISCO codes and the translation documents from ISCO-88 to ISCO-08 to clarify any questions in relation to the new scheme.

In their national options, countries may also have needed to pre-code responses to some items before data from the questionnaire were entered into the software.

DATA ENTRY, DATA CHECKING AND FILE SUBMISSION

Data entry

The international contractor provided participating countries with *KeyQuest* data entry software. *KeyQuest* contained the database structures for all of the booklets, questionnaires and tracking forms used in the Main Survey. Variables could be added or deleted as needed for national options. Approved adaptations to response categories could also be accommodated. Student response data were entered directly from the test booklets and questionnaires. Information regarding the participation of students, recorded by the School Co-ordinator and Test Administrator on the session attendance form, was entered directly into *KeyQuest*. Several questions from the session report form, such as the timing of the session, were also entered into *KeyQuest*.

KeyQuest performed validation checks as data were entered. Importing facilities were also available if data had already been entered into text files, but it was strongly recommended that data be entered directly into *KeyQuest* to take advantage of its PISA-specific features. A *Data Management Manual* provided complete instructions specific to the Main Survey regarding data entry, data management and validity checks.

Data checking and submission

NPMs were responsible for ensuring that checks of the quality of their country's data were made before the data files were submitted to the international contractor. These checks were explained in detail in the *Data Management Manual*, and could be simply applied using the *KeyQuest* software. The checking procedures required that the list of sampled schools and the session attendance form for each school were already accurately completed and entered into *KeyQuest*. Any errors had to be corrected before the data were submitted. Copies of the cleaning reports were to be submitted together with the data files. More details on the cleaning steps are provided in Chapter 10.

Data were submitted through the ACER *KeyQuest* database.

After data were submitted

NPMs were required to designate a data manager who would work actively with the international contractor's data processing centre at ACER during the international data cleaning process. Responses to requests for information by the processing centre were required within three working days of the request.

THE MAIN SURVEY REVIEW

NPMs were required to complete a structured review of their Main Survey operations. The review was an opportunity to provide feedback to the international contractor on the various aspects of the implementation of PISA, and to provide suggestions for areas that could be improved. It also provided an opportunity for the NPM to formally document aspects such as the operational structure of the National Centre, the security measures that were implemented, and the use of contractors for particular activities and so on.

The Main Survey review was submitted to the international contractor four weeks after the submission of the national database.

Notes

1. The document *Procedures for coding paper-based constructed-response items MS12* is available at *www.oecd.org/pisa*.

2. Coding design options are discussed in the document *Procedures for coding paper-based constructed-response items MS12* available at *www.oecd.org/pisa*.

Reference

International Labour Organization (ILO) (2007), "ILO plans to support implementation of ISCO-08 in national and regional activities", Paper for discussion by the United Nations Expert Group on International Economic and Social Classifications, New York, April 16-18, 2007.

7

PISA Quality Monitoring

PISA data collection activities are undertaken in accordance with strict quality assurance procedures. The quality assurance that ensures the PISA 2012 data are fit for use consists of two components. The first is to develop and document procedures for data collection and the second is to monitor and record the implementation of those procedures. This chapter considers the second part of the process – monitoring quality.

While the aim of quality control is to establish effective and efficient procedures and guide implementation process, quality monitoring activities are set to observe and record any deviations from those agreed procedures during the implementation of the survey.

- Field Trial and Main Survey review
- Final Optical Check
- National Centre Quality Monitor (NCQM) visits and consultations
- PISA Quality Monitor (PQM) visits
- Test Administration
- Post Final Optical Check

FIELD TRIAL AND MAIN SURVEY REVIEW

After the implementation of the Field Trial and the Main Survey, National Project Managers (NPM) were given the opportunity to review and provide feedback to the international contractor on all aspects of the field operations. This information is used to guide future implementations of the assessment.

The Field Trial and Main Survey reviews were organised around all aspects outlined in the NPM manual:

- use of key documents and processes: use a rating system to review NPMs' level of satisfaction with the clarity of key documents and manuals;
- communication with the international contractor;
- review the usefulness of the MyPISA website; as well as using a rating system to review the communication by activity;
- implementation of national and international options: confirm if National Centre had executed any national and international options as agreed;
- review the outcomes of and process for provision of national feedback on proposed test items;
- security arrangements: review security arrangements to confirm if they had been implemented;
- sampling plan: confirm if the PISA Field Trial test was implemented as agreed in the sampling plan;
- translation/adaptation/verification: review the translation, adaptation and verification processes to see if they were implemented in accordance with PISA technical standards and to a satisfactory level;
- archiving of materials: confirm if the National Centre had archived the test materials in accordance with the technical standards;
- printing: review the print quality agreement process;
- Test Administration: review Test Administrators training processes and Test Administration procedures;
- quality assurance: review the Field Trial PISA Quality Monitoring activity at national level, as well as the PQM activity during Main Survey at international level;
- coding: review coder training procedures, coding procedures, coding designs and the time required for coding; and
- data management: review the data management processes, including student sampling, database adaptation, data entry, coding of occupational categories, validity reports and data submission.

FINAL OPTICAL CHECK

Before printing assessment materials in each participating country, NPMs electronically submit their final version of the test booklets to the international contractor for a Final Optical Check (FOC). The FOC is undertaken by the international contractor's verifiers and involves a page-by-page inspection of test booklets and questionnaire forms with regard to correct implementation of agreed adaptations, correct item allocation to test forms, layout, page numbering, item numbering, graphic elements, item codes, footers and so on (see Chapter 5).

Any error found during the FOC is recorded and forwarded to National Centres for correction.

NATIONAL CENTRE QUALITY MONITOR (NCQM) VISITS AND CONSULTATIONS

Most countries participating in PISA 2012 had already been involved in the PISA assessment, so only new countries without international assessment experience or countries in which problems had arisen were visited. This resulted in visits to three countries. Visits were usually carried out during the Field Trial period so that preventive and corrective action could be taken if any potential problems were detected.

During the visits, the NCQM conducted a face-to-face interview with the NPM or a representative from the National Centre. Any potential problems identified by the NCQM were forwarded to the relevant international contractor expert for appropriate action. A collated response to all problems identified was sent back to the visited National Centre after the visit.

The NCQMs have comprehensive knowledge and extensive experience regarding PISA operations. Each NCQM was trained and provided with the National Centre's project implementation data in great detail. Prior to each visit, NCQMs studied the national materials in order to be familiar with country-specific information during the interview with NPMs.

The purpose of this interview is twofold. Firstly, it allows staff members of the international contractor to become familiar with the operations of PISA in national context, as well as any specific challenges 'new countries' may be facing in national contexts. Secondly, it provides National Centre staff with the opportunity to ask questions or receive clarification about any aspect of the survey.

The NCQM interview schedule is a list of areas that was prepared for the international contractor representatives to lead the interview in a structured way, so that the outcomes of the NCQM site visit could be recorded systematically and consistently across countries. This interview schedule covers the following areas:

- General organisation of PISA in each country
- Sampling
- Adaptation, translation and printing of tests, questionnaires and operational materials
- Dispatch of materials and Test Administration
- Security and checking back of materials
- Cognitive item coding
- Data management and submission

As well as more formal NCQM visits, a large number of consultations meetings took place between senior staff of the international contractor and NPMs or other representatives of National Centres, in the context of NPM meetings. An extensive schedule of consultation meetings was developed prior to each meeting, and the consultations provided the opportunity for detailed discussion on a wide variety of PISA implementation matters on which additional advice or support was sought by the National Centre.

In addition, the international contractor was in constant communication with all countries through email and via the MyPISA website.

PISA QUALITY MONITOR (PQM) VISITS

The International Contractor appoints a single PISA Quality Monitoring co-ordinator. This person is familiar with all aspects of the implementation of PISA. The PQM co-ordinator's role involves directly recruiting the PQMs in each of the countries, organising their training, approving the list of schools to visit and collecting information from the PQM visits.

PQMs are individuals employed by the international contractor and located in participating countries. They visit a sample of schools to observe Test Administration and to record the implementation of the documented field operations procedures in the Main Survey. Typically, one or two PQMs were hired for each country and they visit a total of 7 or 8 schools in each country.

All PQMs are nominated by the NPMs through a formal process of submission of nominations to ACER. Based upon the NPM nominations, which are accompanied by candidate resumes, the PISA Quality Monitoring co-ordinator selects PQMs who are independent from the National Centre, knowledgeable in testing procedures or with a background in education and research, and able to fluently communicate in English. Where the resume does not match the selection criteria, further information or an alternate nomination is sought.

The *PQM Manual*, PQM self-training package, other operational manuals and copies of data collection sheets were made available to all PQMs upon receipt of their signed confidentiality agreement via emails and post. The PQMs were also given access to a designated PQM web page on the MyPISA website from which they could download materials and information. All PQMs were self-trained using the PQM training PowerPoint®, which has an embedded soundtrack. At the same time, the PQM co-ordinator provided support and addressed any issues or concerns via email. The PQMs and the PQM co-ordinator collaborated to develop a schedule of Test Administration site visits to ensure that a range of schools was covered and that the schedule of visits was both economically and practically feasible. ACER paid the expenses and fees directly to each PQM. In most countries, seven visits to schools were carried out.

The PISA School Co-ordinator in each school is responsible for providing a link between the PISA National Project Manager and the school, its students, teachers and principal. The School Co-ordinator provides the list of 15-year-olds from whom the school's random sample is chosen and organises a suitable venue for the testing. The PQM for each country is supplied with a list of School Co-ordinators by the PISA National Project Manager and makes contact with the relevant School Co-ordinators directly after a school has been selected for a PQM visit.

The majority of school visits were unannounced to the Test Administrator. However, in some countries it was not possible to do so when the school associate model was used, where the Test Administrator and the School Co-ordinator are the same person.

A PQM data collection form was developed for PQMs to systematically record their observations during each school visit. The data collection form covers the following areas:

- Preparation for the assessment
- Conducting the assessment
- General questions concerning the assessment
- Interview with the School Co-ordinator

TEST ADMINISTRATION

Test Administrators record all key test session information using a test session report. This report provides detailed data on Test Administration, including:

- Session date and timing
- Position of the Test Administrator
- Conduct of the students
- Testing environment

This information was used to check that the implementation in each school was in accordance with the PISA Technical Standards. The information was also called upon if a country's results showed, for example, a greater degree of country-item interaction.

DATA ADJUDICATION

All quality assurance data collected throughout the cycle are entered and collated in a central data adjudication database. Comprehensive reports are then generated for the Technical Advisory Group (TAG) for consideration during the data adjudication process (see Chapter 14).

The TAG experts use the consolidated quality-monitoring reports from the central data adjudication database to make country-by-country evaluations on the quality of field operations, printing, translation, school and student sampling, and coding. The final reports by TAG experts are then used for the purpose of data adjudication.

POST FINAL OPTICAL CHECK

After both the Field Trial and Main Survey, international contractor staff carried out a thorough checking procedure on all the hard copies of the National Centre test booklets that had been submitted to the international contractor for archiving purpose. The checking was carried out by comparing the National Centres' submitted booklets and the source version of the test booklets that were released by the international contractor, as well as checking issues that were identified during the FOC process to see how well the suggested changes were implemented and to what extent.

Findings were recorded, in particular observed errors were recorded in the data adjudication database.

8

Survey Weighting and the Calculation of Sampling Variance

Survey weights are required to facilitate analysis of PISA data, calculation of appropriate estimates of sampling error and making valid estimates and inferences of the population. The international contractor calculated survey weights for all assessed, ineligible and excluded students, and provided variables in the database that permit users to make approximately unbiased estimates of standard errors, conduct significance tests and create confidence intervals appropriately, given the complex sample design for PISA in each individual participating country.

SURVEY WEIGHTING

While the students included in the final PISA sample for a given country were chosen randomly, the selection probabilities of the students vary. Survey weights must therefore be incorporated into the analysis to ensure that each sampled student represents the appropriate number of students in the full PISA population.

There are several reasons why the survey weights are not the same for all students in a given country:

- A school sample design may intentionally over- or under-sample certain sectors of the school population: in the former case, so that they could be effectively analysed separately for national purposes, such as a relatively small but politically important province or region, or a sub-population using a particular language of instruction; and in the latter case, for reasons of cost, or other practical considerations, such as very small or geographically remote schools.[1]
- Information about school size available at the time of sampling may not have been completely accurate. If a school was expected to be large, the selection probability was based on the assumption that only a sample of students would be selected from the school for participation in PISA. But if the school turned out to be small, all students would have to be included. In this scenario, the students would have a higher probability of selection in the sample than planned, making their inclusion probabilities higher than those of most other students in the sample. Conversely, if a school assumed to be small actually was large, the students included in the sample would have smaller selection probabilities than others.
- School non-response, where no replacement school participated, may have occurred, leading to the under-representation of students from that kind of school, unless weighting adjustments were made. It is also possible that only part of the PISA-eligible population in a school (such as those 15-year-old students in a particular grade) were represented by its student sample, which also requires weighting to compensate for the missing data from the omitted grades.
- Student non-response, within participating schools, occurred to varying extents. Sampled students who were PISA-eligible and not excluded, but did not participate in the assessment for reasons such as absences or refusals, will be under-represented in the data unless weighting adjustments were made.
- Trimming the survey weights to prevent undue influence of a relatively small subset of the school or student sample might have been necessary if a small group of students would otherwise have much larger weights than the remaining students in the country. Such large survey weights can lead to estimates with large sampling errors and inappropriate representations in the national estimates. Trimming survey weights introduces a small bias into estimates but greatly reduces standard errors (Kish, 1992).
- In countries that participated in the financial literacy study, additional students were selected in the schools eligible for the financial literacy assessment. Since the financial literacy sample was also designed to represent the full PISA student population, the weights for the sampled students were adjusted to account for this. Different adjustment factors applied to each student's weight, depending on whether the student was sampled for financial literacy or not.

The procedures used to derive the survey weights for PISA reflect the standards of best practice for analysing complex survey data, and the procedures used by the world's major statistical agencies. The same procedures were used in other international studies of educational achievement such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Studies (PIRLS), which were all implemented by the International Association for the Evaluation of Educational Achievement (IEA). The underlying statistical theory for the analysis of survey data can be found in Cochran (1977), Lohr (2010) and Särndal, Swensson and Wretman (1992).

Weights are applied to student-level data for analysis. The weight, W_{ij}, for student j in school i consists of two base weights, the school base weight and the within-school base weight, and five adjustment factors, and can be expressed as:

8.1

$$W_{ij} = t_{2ij} f_{1i} f_{2ij} f^{A}_{1ij} t_{1i} w_{2ij} w_{1i}$$

where:

w_{1i}, the school base weight, is given as the reciprocal of the probability of inclusion of school i into the sample;

w_{2ij}, the within-school base weight, is given as the reciprocal of the probability of selection of student j from within the selected school i;

f_{1i} is an adjustment factor to compensate for non-participation by other schools that are somewhat similar in nature to school i (not already compensated for by the participation of replacement schools);

f^{A}_{1ij} is an adjustment factor to compensate for schools in some participating countries where only 15-year-old students who were enrolled in the modal grade for 15-year-old students were included in the assessment;

f_{2ij} is an adjustment factor to compensate for non-participation by students within the same school non-response cell and explicit stratum, and, where permitted by the sample size, within the same high/low grade and gender categories;

t_{1i} is a school base weight trimming factor, used to reduce unexpectedly large values of w_{1i}; and

t_{2ij} is a final student weight trimming factor, used to reduce the weights of students with exceptionally large values for the product of all the preceding weight components.

The school base weight

The term w_{1i} is referred to as the school base weight. For the systematic sampling with probability proportional-to-size method used in sampling schools for PISA, this weight is given as:

8.2

$$w_{1i} = \begin{cases} {}^{I_g}\!/_{MOS_i} & \text{if } < MOS_i < I_g \\ 1 & \text{otherwise} \end{cases}$$

The term MOS_i denotes the Measure of Size given to each school on the sampling frame.

Despite country variations, MOS_i was usually equal to the estimated number of 15-year-old students in the school, if it was greater than the predetermined target cluster size (*TCS*), which was 35 students for most countries that did not participate in the financial literacy study, and 43 for most countries that did. If small schools were under-sampled without the use of an explicit stratum for small school, then if the enrolment of 15-year-old students was less than the *TCS*, but greater than $TCS/2$, $MOS_i = TCS$. If the enrolment was between 3 and $TCS/2$, $MOS_i = TCS/2$ and if the enrolment was 1 or 2, $MOS_i = TCS/4$. These different values of the measure of size are intended to minimise the impact of small schools on the variation of the weights.

The term I_g denotes the sampling interval used within the explicit sampling stratum g that contains school i and is calculated as the total of the MOS_i values for all schools in stratum g, divided by the school sample size for that stratum.

Thus, if school i was estimated to have one hundred 15-year-old students at the time of sample selection, $MOS_i = 100$. If the country had a single explicit stratum ($g=1$) and the total of the MOS_i values over all schools was 150 000 students, with a school sample size of 150, then the sampling interval, I_1 = 150 000/150 = 1 000, for school i (and others in the sample), giving a school base weight of $w_{1i} = 1000/100 = 10.0$. Thus, the school can be thought of as representing about ten schools in the population. In this example, any school with 1 000 or more 15-year-old students would be included in the sample with certainty, with a base weight of $w_{1i} = 1$ as the MOS_i is larger than the sampling interval.

The school base weight trimming factor

Once school base weights were established for each sampled school in the country, verifications were made separately within each explicit sampling stratum to determine if the school base weights required trimming. The school trimming factor t_{1i}, is the ratio of the trimmed to the untrimmed school base weight, and for most schools (and therefore most students in the sample) is equal to 1.0000.

The school-level trimming adjustment was applied to schools that turned out to be much larger than was assumed at the time of school sampling. Schools were flagged where the actual 15-year-old student enrolment exceeded $3 \times maximum\ (TCS, MOS_i)$. For example, if the *TCS* was 35 students, then a school flagged for trimming had more than 105 (=3 x 35) PISA-eligible students, and more than three times as many students as was indicated on the school sampling frame. Because the student sample size was set at *TCS* regardless of the actual enrolment, the student sampling rate was much lower than anticipated during the school sampling. This meant that the weights for the sampled students in these schools would have been more than three times greater than anticipated when the school sample was selected. These schools had their school base weights trimmed by having MOS_i replaced by $3 \times maximum\ (TCS, MOS_i)$ in the school base weight formula.

The within-school base weight

The term w_{2ij} is referred to as the within-school base weight. With the PISA procedure for sampling students, w_{2ij} did not vary across students within a particular school *i*. That is, all of the students within the same school had the same probability of selection for participation in PISA. This weight is given as:

8.3

$$w_{2ij} = {}^{enr_i}\!\big/_{sam_i}$$

where enr_i is the actual enrolment of 15-year-old students in the school on the day of the assessment (and so, in general, is somewhat different from the MOS_i), and sam_i is the sample size within school *i*. It follows that if all PISA-eligible students from the school were selected, then $w_{2ij} = 1$ for all eligible students in the school. For all other cases $w_{2ij} > 1$ as the selected student represents other students in the school besides themselves.

In the case of the grade sampling option, for direct sampled grade students, the sampling interval for the extra grade students was the same as that for the PISA students. Therefore, countries with extra direct-sampled grade students (Iceland and some of the grade students in Switzerland) have the same within school student weights for the extra grade students as those for PISA-eligible students from the same school. For Switzerland's other grade sampled students, these had weights of 1. For Slovenia, a separate sample size was specified for the non-PISA grade students and so their weights differed from those of the PISA students in the same school.

Additional weight components were needed for the grade students in Chile and Germany. For these two countries, the extra weight component consisted of the class weight for the selected class(es) (all students were selected into the grade sample in the selected class[es]). In these two countries, the extra weight component resulted in the necessity of a second weighting stream for the extra grade students.

The school non-response adjustment

In order to adjust for the fact that those schools that declined to participate, and were not replaced by a replacement school, were not in general typical of the schools in the sample as a whole, school-level non-response adjustments were made. Several groups of somewhat similar schools were formed within a country, and within each group the weights of the responding schools were adjusted to compensate for the missing schools and their students.

The compositions of the non-response groups varied from country to country, but were based on cross-classifying the explicit and implicit stratification variables used at the time of school sample selection. Usually, about 10 to 15 such groups were formed within a given country depending upon school distribution with respect to stratification variables. If a country provided no implicit stratification variables, schools were divided into three roughly equal groups, within each explicit stratum, based on their enrolment size. It was desirable to ensure that each group had at least six participating schools, as small groups could lead to unstable weight adjustments, which in turn would inflate the sampling variances. Adjustments greater than 2.0 were also flagged for review, as they could have caused increased variability in the weights

and would have led to an increase in sampling variances. It was not necessary to collapse cells where all schools participated, as the school non-response adjustment factor was 1.0 regardless of whether cells were collapsed or not. However, such cells were sometimes collapsed to ensure that enough responding students would be available for the student non-response adjustments in a later weighting step. In either of these situations, cells were generally collapsed over the last implicit stratification variable(s) until the violations no longer existed. In participating countries with very high overall levels of school non-response after school replacement, the requirement for school non-response adjustment factors to all be below 2.0 was waived.

Within the school non-response adjustment group containing school *i*, the non-response adjustment factor was calculated as:

8.4

$$f_{1i} = \frac{\sum_{k \in \Omega(i)} w_{1k} enr(k)}{\sum_{k \in \Gamma(i)} w_{1k} enr(k)}$$

where the sum in the denominator is over Γ(*i*), which are the schools within the group (originals and replacements) that participated, while the sum in the numerator is over Ω(*i*), which are those same schools, plus the original sample schools that refused and were not replaced. The numerator estimates the population of 15-year-old students in the group, while the denominator gives the size of the population of 15-year-old students directly represented by participating schools. The school non-response adjustment factor ensures that participating schools are weighted to represent all students in the group. If a school did not participate because it had no PISA-eligible students enrolled, no adjustment was necessary since this was considered neither non-response nor under-coverage.

Figure 8.1 shows the number of school non-response classes that were formed for each country/economy and the variables that were used to create the cells.

The grade non-response adjustment

Because of perceived administrative inconvenience, individual schools may occasionally agree to participate in PISA but require that participation be restricted to 15-year-old students in the modal grade for 15-year-old students, rather than all 15-year-old students. Since the modal grade generally includes the majority of the population to be covered, such schools may be accepted as participants rather than have the school refuse to participate entirely. For the part of the 15-year-old population in the modal grade, these schools are respondents, while for the rest of the grades in the school with 15-year-old students, such a school is a refusal. To account for this, a special non-response adjustment can be calculated at the school level for students not in the modal grade (and is automatically 1.0 for all students in the modal grade). No countries had this type of non-response for PISA 2012, so the weight adjustment for grade non-response was automatically 1.0 for all students in both the modal and non-modal grades, and therefore did not affect the final weights.

If the weight adjustment for grade non-response had been needed (as it was in earlier cycles of PISA in a few countries), it would have been calculated as follows:

Within the same non-response adjustment groups used for creating school non-response adjustment factors, the grade non-response adjustment factor for all students in school *i*, f_{1i}^{A}, is given as:

8.5

$$f_{1i}^{A} = \begin{cases} \sum_{k \in C(i)} w_{1k} enra(k) \Big/ \sum_{k \in B(i)} w_{1k} enra(k) & \text{for students not in the modal grade} \\ 1 & \text{otherwise} \end{cases}$$

The variable *enra*(*k*) is the approximate number of 15-year-old students in school *k* but not in the modal grade. The set B(*i*) is all schools that participated for all eligible grades (from within the non-response adjustment group with school (*i*)), while the set C(*i*) includes these schools and those that only participated for the modal responding grade.

■ Figure 8.1 [Part 1/2] ■

Non-response classes

Country/ Economy	Implicit stratification variables	Number of original cells (2012)	Number of final cells (2012)
Albania	ISCED2/Mixed (2)	18	8
Argentina	Funding (2); Education type (3); Education level (9); Urbanicity (2); Secular/Religious (2)	83	14
Australia	Geographic Zone (3); School Gender Composition (3); School Socio-economic Level (6); Numeracy Achievement Level (6); ISCED Level (3)	455	84
Austria	School Type (4); Region (9); Percentage of Girls (5)	191	22
Belgium	Grade Repetition – Flemish Community and French Community (5), German Community (1); Percentage of Girls – Flemish Community and French Community (4), German Community (1); School Type – French Community (4), German Community and Flemish Community (1)	224	36
Brazil	Admin (3); DHI Quintiles (6); ISCED level (4); Urbanicity (2)	420	118
Bulgaria	Type of School (8); Size of Settlement (5); Funding (3)	131	28
Canada	Urbanicity (3); Funding (2); ISCED Level (4)	194	57
Chile	% Girls (6); Urbanicity (2); Region (4)	156	22
Colombia	Urbanicity (2); Funding (2); Weekend school or not (2); Gender (5); ISCED Programme Orientation (4)	113	26
Costa Rica	Programme (2); Urbanicity (2); Shift (2); Region (27); ISCED Level (3)	93	11
Croatia	Gender (3); Urbanicity (3); Region (6)	78	27
Cyprus[1, 2]	Language (2); ISCED Level (3)	14	10
Czech Republic	School Size (3); Region for Programmes 3, 4, 5, 6 (15); School Gender Composition (3)	194	39
Denmark	School Type (8); ISCED Level (4); Urbanicity (6); Region (6)	164	42
Estonia	School Type (3); Urbanicity (2); County (15); Funding (2)	81	24
Finland	School Type (7)	52	20
France	School Type for small school strata (4); Funding (2)	19	8
Germany	State for other schools (17); School Type (6)	79	34
Greece	School Type (3); Funding (2)	44	15
Hong Kong-China	Student Academic Intake (4)	11	8
Hungary	Region (7); Mathematics Performance (6)	122	31
Iceland	Urbanicity (2); ISCED Level (2)	41	14
Indonesia	Province (32); Funding (2); School Type and Level (5); National Exam Result (3)	148	27
Ireland	Socio-Economic Status Category (5); School Gender Composition Category (5)	93	25
Israel	ISCED Level (4); Group Size (3); SES (4); District (3)	67	17
Italy	Funding (2)	152	69
Japan	Levels of proportion of students taking University/College Entrance Exams (4)	16	13
Jordan	Urbanicity (2); Gender (3); Level (2); Shift (2)	52	25
Kazakhstan	Urbanicity (2); ISCED Level (3); ISCED Programme Orientation (2); Funding (2)	128	35
Korea	Urbanicity Level (3); School Gender Composition (3)	24	13
Latvia	School Type/Level (5)	18	9
Liechtenstein	Funding (2)	2	2
Lithuania	Funding (2)	21	12
Luxembourg	School Gender Composition (3)	8	8
Macao-China	Gender (3); School Orientation (2); ISCED Level (2)	19	13
Malaysia	School Type (16); Urbanicity (2); State (16); Gender (3); ISCED Level (2)	61	17
Mexico	School Level (2); School Programme (7); Funding (2); Urbanicity (2)	610	124
Montenegro	Gender (3)	17	15
Netherlands	Programme Category (7)	14	6
New Zealand	School Decile (4); Funding (2); School Gender Composition (3); Urbanicity (2)	37	16
Norway	None	8	4
Peru	Region (26); Gender (3); School Type (7)	104	27
Poland	School Sub-type (2); Funding (2); Locality (4); Gender Composition (3)	36	7
Portugal	ISCED Level (3); Funding (2); Urbanicity (3)	101	31
Qatar	Gender (3); Language (2); Level (5); Funding (2); Programme Orientation (3)	42	13
Romania	Language (2); Urbanicity (2); LIC Type (3)	10	5
Russian Federation	Location/Urbanicity (9); School Type (8); School Sub-type (5)	193	43
Serbia	Region (5); Programme (7)	38	18
Shanghai-China	Urbanicity (2); Funding (2); Vocational School Type (4)	19	17
Singapore	Gender (3)	6	4
Slovak Republic	Sub-Type (6); Language (3); Grade Repetition Level (25); Exam (11)	96	34
Slovenia	Location/Urbanicity (5); Gender (3)	146	43

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

■ Figure 8.1 [Part 2/2] ■

Non-response classes

Country/ Economy	Implicit stratification variables	Number of original cells (2012)	Number of final cells (2012)
Spain	None	129	105
Sweden	Geographic LAN (22); Responsible Authority (4); Level of Immigrants (5); Income Quartiles (5)	114	33
Switzerland	School Type (28); Canton (26)	144	36
Chinese Taipei	County/City area (22); School Gender (3)	125	41
Thailand	Region (9); Urbanicity (2); Gender (3)	118	27
Tunisia	ISCED Level (3); Funding (2); % Repeaters (3)	85	23
Turkey	School Type (18); Gender (3); Urbanicity (2); Funding (2)	128	27
United Arab Emirates	School Level (3); School Gender (3)	128	59
United Kingdom	Gender (3); School Performance – England and Wales (6), Northern Ireland (1); Local Authority – England (151), Wales (22), Northern Ireland (1); Area Type – Scotland (6)	339	66
United States	Grade Span (5); Urbanicity (4); Minority Status (2); Gender (3); State (51)	223	32
Uruguay	Location/Urbanicity (4); Gender (4)	33	16
Viet Nam	Economic Region (8); Province (63); School Type (6); Study Commitment (2)	142	28

This procedure gives, for each school, a single grade non-response adjustment factor that depends upon its non-response adjustment class. Each individual student has this factor applied to the weight if he/she did not belong to the modal grade, and 1.0 if belonging to the modal grade. In general, this factor is not the same for all students within the same school when a country has some grade non-response.

The within school non-response adjustment

The final level of non-response adjustment was at the student level. Student non-response adjustment cells were created by forming four cells within each school, by cross-classifying gender with grade, dichotomised into "high" and "low" categories. The definition as to which grades were high and which were low varied across explicit strata, with the aim of making the two groups as equal in size as possible. In general the cells formed in this way were too small for the formation of stable nonresponse adjustment factors (sometimes such cells even contained no responding students). Thus cells were collapsed to create final student non-response adjustment cells. Initially the collapsing was across schools, within school non-response adjustment classes. Then as necessary either grade or gender was collapsed. The student non-response adjustment f_{2i} was calculated as:

8.6

$$f_{2i} = \frac{\sum_{k \in X(i)} f_{1i} w_{1i} w_{2ik}}{\sum_{k \in \Delta(i)} f_{1i} w_{1i} w_{2ik}}$$

where

$\Delta(i)$ is all assessed students in the final student non-response adjustment cell; and,

$X(i)$ is all assessed students in the final student non-response adjustment cell, plus all other students who should have been assessed (i.e. who were absent, but not excluded or ineligible).

As mentioned, the high and low grade categories within each explicit stratum in each country were defined so as to each contain a substantial proportion of the PISA population.

In most cases, this student non-response factor reduces to the ratio of the number of students who should have been assessed to the number who were assessed. In some cases where it was necessary to collapse cells together, and then apply the more complex formula shown above was required. Additionally, an adjustment factor greater than 2.0 was not allowed for the same reasons noted under school non-response adjustments. If this occurred, the cell with the large adjustment was collapsed with the closest cell within grade and gender combinations in the same school non-response cell.

Some schools in some countries had extremely low student response levels. In these cases it was determined that the small sample of assessed students within the school was potentially too biased as a representation of the school to be included in the final PISA dataset. For any school where the student response rate was below 25%, the school was treated

as a non-respondent, and its student data were removed. In schools with between 25 and 50% student response, the student non-response adjustment described above would have resulted in an adjustment factor of between 2.0 and 4.0, and so the grade-gender cells of these schools were collapsed with others to create student non-response adjustments.[2]

For countries with extra direct grade sampled students (Iceland, Slovenia and Switzerland), care was taken to ensure that student non-response cells were formed separately for PISA students and the extra non-PISA grade students. No procedural changes were needed for Chile and Germany since a separate weighting stream was needed for the grade students.

Trimming the student weights

This final trimming check was used to detect individual student weights that were unusually large compared to those of other students within the same explicit stratum. The sample design was intended to give all students from within the same explicit stratum an equal probability of selection and therefore equal weight, in the absence of school and student non-response. As already noted, poor prior information about the number of eligible students in each school could lead to substantial violations of this equal weighting principle. Moreover, school, grade, and student non-response adjustments, and, occasionally, inappropriate student sampling could, in a few cases, accumulate to give a few students in the data relatively large weights, which adds considerably to the sampling variance. The weights of individual students were therefore reviewed, and where the weight was more than four times the median weight of students from the same explicit sampling stratum, it was trimmed to be equal to four times the median weight for that explicit stratum.

The student trimming factor, t_{2ij}, is equal to the ratio of the final student weight to the student weight adjusted for student non-response, and therefore equal to 1.0 for the great majority of students. The final weight variable on the data file is the final student weight that incorporates any student-level trimming. As in PISA 2000, PISA 2003, PISA 2006 and PISA 2009 minimal trimming was required at either the school or the student levels.

The financial literacy adjustment factor

The financial literacy weighting adjustment factor was applied to all students in the schools sampled for the financial literacy study. Despite difference in *TCS* values and number of financial literacy sampled students, the factors were the same for almost all countries. The financial literacy booklet was applied at a rate of *43/8*, which then became the adjustment factor for students who received the financial literacy booklet. For the remaining students, the factor was *43/35*, the rate at which non-financial literacy booklets were applied. Alternative factors were used for whole schools using the "une heure" booklet (UH) to reflect the slightly different rate at which the financial literacy UH booklet was applied in those schools (16/3 for students receiving a financial literacy booklet and 16/13 for students not receiving a financial literacy booklet) (see Chapter 2 for further details on the UH booklet).

Weighting for computer-based assessment

No non-response adjustments were made for schools or students sampled for computer-based assessment (CBA) which did not participate. Since CBA was being treated as a minor domain like mathematics and reading, absent CBA students were treated in the same manner as a student not assigned a booklet containing items in the mathematics or reading domain. Plausible values were generated for these CBA students, as well as for all other students who had not been subsampled for CBA. For most countries, CBA final sample sizes are therefore identical to sample sizes of paper-based tests. Sample weights and replicate can be used without any modification.

The school subsampling for CBA for Brazil, Italy and Spain needed to be accounted for in weighting through an additional weight component. Thus, schools subsampled for CBA for Brazil, Italy and Spain had their own weighting stream, separate from the weighting stream for the large national samples in these countries. Once in their own weighting stream, weighting procedures for these CBA subsampled schools and students were the same as the weighting procedures used for all countries.

CALCULATING SAMPLING VARIANCE

A replication methodology was employed to estimate the sampling variances of PISA parameter estimates. This methodology accounted for the variance in estimates due to the sampling of schools and students. Additional variance due to the use of plausible values from the posterior distributions of scaled scores was captured separately as measurement error. Computationally the calculation of these two components could be carried out in a single program, such as *WesVar 5.1* (Westat, 2007). SPSS and SAS macros were also developed. For further detail, see *PISA Data Analysis Manual* (OECD, 2009).

The Balanced Repeated Replication variance estimator

The approach used for calculating sampling variances for PISA estimates is known as Balanced Repeated Replication (BRR), or balanced half-samples; the particular variant known as Fay's method was used. This method is similar in nature to the jackknife method used in other international studies of educational achievement, such as TIMSS, and it is well documented in the survey sampling literature (see Rust, 1985; Rust and Rao, 1996; Shao, 1996; Wolter, 2007). The major advantage of the BRR method over the jackknife method is that the jackknife is not fully appropriate for use with non-differentiable functions of the survey data, most noticeably quantiles, for which it does not provide a statistically consistent estimator of variance. This means that, depending upon the sample design, the variance estimator can be unstable, and despite empirical evidence that it can behave well in a PISA-like design, theory is lacking. In contrast the BRR method does not have this theoretical flaw. The standard BRR procedure can become unstable when used to analyse sparse population subgroups, but Fay's method overcomes this difficulty, and is well justified in the literature (Judkins, 1990).

The BRR method was implemented for a country where the student sample was selected from a sample of schools, rather than all schools, as follows:

- Schools were paired on the basis of the explicit and implicit stratification and frame ordering used in sampling. The pairs were originally sampled schools, except for participating replacement schools that took the place of an original school. For an odd number of schools within a stratum, a triple was formed consisting of the last three schools on the sorted list.
- Pairs were numbered sequentially, 1 to H, with pair number denoted by the subscript h. Other studies and the literature refer to such pairs as variance strata or zones, or pseudo-strata.
- Within each variance stratum, one school was randomly numbered as 1, the other as 2 (and the third as 3, in a triple), which defined the variance unit of the school. Subscript j refers to this numbering.
- These variance strata and variance units (1, 2, 3) assigned at school level were attached to the data for the sampled students within the corresponding school.
- Let the estimate of a given statistic from the full student sample be denoted as X^*. This was calculated using the full sample weights.
- A set of 80 replicate estimates, X_t^* (where t runs from 1 to 80), was created. Each of these replicate estimates was formed by multiplying the survey weights from one of the two schools in each stratum by 1.5, and the weights from the remaining schools by 0.5. The determination as to which schools received inflated weights, and which received deflated weights, was carried out in a systematic fashion, based on the entries in a Hadamard matrix of order 80. A Hadamard matrix contains entries that are +1 and –1 in value, and has the property that the matrix, multiplied by its transpose, gives the identity matrix of order 80, multiplied by a factor of 80. Details concerning Hadamard matrices are given in Wolter (2007).
- In cases where there were three units in a triple, either one of the schools (designated at random) received a factor of 1.7071 for a given replicate, with the other two schools receiving factors of 0.6464, or else the one school received a factor of 0.2929 and the other two schools received factors of 1.3536. The explanation of how these particular factors came to be used is explained in Appendix 12 of the *PISA 2000 Technical Report* (Adams and Wu, 2002).
- To use a Hadamard matrix of order 80 requires that there be no more than 80 variance strata within a country, or else that some combining of variance strata be carried out prior to assigning the replication factors via the Hadamard matrix. The combining of variance strata does not cause bias in variance estimation, provided that it is carried out in such a way that the assignment of variance units is independent from one stratum to another within strata that are combined. That is, the assignment of variance units must be completed before the combining of variance strata takes place, and this approach was used for PISA.
- The reliability of variance estimates for important population subgroups is enhanced if any combining of variance strata that is required is conducted by combining variance strata from different subgroups. Thus in PISA, variance strata that were combined were selected from different explicit sampling strata and also, to the extent possible, from different implicit sampling strata.
- In some countries, it was not the case that the entire sample was a two-stage design, of first sampling schools and then sampling students within schools. In some countries/economies, for part of the sample (and for the entire samples for Cyprus,[3] Iceland, Liechtenstein, Luxembourg, Macao-China and Qatar), schools were included with certainty into the sampling, so that only a single stage of student sampling was carried out for this part of the sample. In these cases

instead of pairing schools, pairs of individual students were formed from within the same school (and if the school had an odd number of sampled students, a triple of students was formed). The procedure of assigning variance units and replicate weight factors was then conducted at the student level, rather than at the school level.

- In contrast, in one country, the Russian Federation, there was a stage of sampling that preceded the selection of schools. Then the procedure for assigning variance strata, variance units and replicate factors was applied at this higher level of sampling. The schools and students then inherited the assignment from the higher-level unit in which they were located.
- Procedural changes were in general not needed in the formation of variance strata for countries with extra direct grade sampled students (Iceland, Slovenia and Switzerland) since the extra grade sample came from the same schools as the PISA students. However, if there were certainty schools in these countries, students within the certainty schools were paired so that PISA non-grade students were together, PISA grade students were together and non-PISA grade students were together. No procedural changes were required for the grade students for Chile and Germany, since a separate weighting stream was needed in these cases.
- The variance estimator is then:

8.7

$$V_{BRR}\left(X^*\right) = 0.05\sum_{t=1}^{80}\left\{\left(X_t^* - X^*\right)^2\right\}$$

The properties of BRR method have been established by demonstrating that it is unbiased and consistent for simple linear estimators (i.e. means from straightforward sample designs), and that it has desirable asymptotic consistency for a wide variety of estimators under complex designs, and through empirical simulation studies.

Reflecting weighting adjustments

This description does not detail one aspect of the implementation of the BRR method. Weights for a given replicate are obtained by applying the adjustment to the weight components that reflect selection probabilities (the school base weight in most cases), and then re-computing the non-response adjustment replicate by replicate.

Implementing this approach required that the international contractor produce a set of replicate weights in addition to the full sample weight. Eighty such replicate weights were needed for each student in the data file. The school and student non-response adjustments had to be repeated for each set of replicate weights.

To estimate sampling errors correctly, the analyst must use the variance estimation formula above, by deriving estimates using the *t*-th set of replicate weights. Because of the weight adjustments (and the presence of occasional triples), this does not mean merely increasing the final full sample weights for half the schools by a factor of 1.5 and decreasing the weights from the remaining schools by a factor of 0.5. Many replicate weights will also be slightly disturbed, beyond these adjustments, as a result of repeating the non-response adjustments separately by replicate.

Formation of variance strata

With the approach described above, all original sampled schools were sorted in stratum order (including refusals, excluded and ineligible schools) and paired. An alternative would have been to pair participating schools only. However, the approach used permits the variance estimator to reflect the impact of non-response adjustments on sampling variance, which the alternative does not. This is unlikely to be a large component of variance in any PISA country, but the procedure gives a more accurate estimate of sampling variance.

Countries and economies where all students were selected for PISA

In Iceland, Liechtenstein, Luxembourg, Macao-China and Qatar, all PISA-eligible students were selected for participation in PISA. It might be unexpected that the PISA data should reflect any sampling variance in these countries/economies, but students have been assigned to variance strata and variance units, and the BRR method does provide a positive estimate of sampling variance for two reasons. First, in each country/economy there was some student non-response, and, in the case of Iceland and Qatar, some school non-response. Not all PISA-eligible students were assessed, giving sampling variance. Second, the intent is to make inference about educational systems and not particular groups of individual students, so it is appropriate that a part of the sampling variance reflect random variation between student populations, even if they were to be subjected to identical educational experiences. This is consistent with the approach that is generally used whenever survey data are used to try to make direct or indirect inference about some underlying system.

Notes

1. Note that this is not the same as excluding certain portions of the school population. This also happened in some cases, but cannot be addressed adequately through the use of survey weights.

2. Chapter 11 describes these schools as being treated as non-respondents for the purpose of response rate calculation, even though their student data were used in the analyses.

3. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

References

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

Cochran, W.G. (1977), *Sampling Techniques*, Third edition, John Wiley and Sons, New York.

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

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

Lohr, S.L. (2010), *Sampling: Design and Analysis*, Second edition, Pacific, Pacific Grove, Duxberry.

OECD (2009), *PISA Data Analysis Manual: SPSS,* Second edition, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264010666-en

OECD (2005), *PISA 2003 Data Analysis Manual: SAS,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264010642-en

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

Rust, K.F. and **J.N.K. Rao** (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., B. Swensson and **J. Wretman** (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

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

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

9

Scaling PISA Cognitive Data

The mixed coefficients multinomial logit model as described by Adams, Wilson and Wang (1997) was used to scale the PISA data, and implemented by *ConQuest* software (Wu, Adams and Wilson, 1997). This chapter presents the model employed, and its application to the analysis of the PISA 2012 data.

THE MIXED COEFFICIENTS MULTINOMIAL LOGIT MODEL

The model applied to PISA is a generalised form of the Rasch model. The model is a mixed coefficients model where items are described by a fixed set of unknown parameters, $\boldsymbol{\xi}$, while the student outcome levels (the latent variable), $\boldsymbol{\theta}$, is a random effect.

Assume that I items are indexed $i = 1,\ldots,I$ with each item admitting $K_i + 1$ response categories indexed $k = 0,1,\ldots,K_i$. Use the vector valued random variable $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ik_i})^T$, where

9.1

$$X_{ij} = \begin{cases} 1 & \text{if response to item } i \text{ is in category } j \\ 0 & \text{otherwise} \end{cases}$$

to indicate the $K_i + 1$ possible responses to item i.

A vector of zeroes denotes a response in category zero, making the zero category a reference category, which is necessary for model identification. Using this as the reference category is arbitrary, and does not affect the generality of the model. The $\mathbf{X}_i$ can also be collected together into the single vector $\mathbf{X}^T = (\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_I^T)$, called the response vector (or pattern). Particular instances of each of these random variables are indicated by their lower case equivalents: x, x_i and x_{ik}.

Items are described through a vector $\boldsymbol{\xi}^T = (\xi_1, \xi_2, \ldots, \xi_p)$, of p parameters. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. A set of design vectors $\mathbf{a}_{ij}$, $(i=1,\ldots,I; j=1,\ldots,K_i)$, each of length p, which can be collected to form a design matrix $\mathbf{A}^T = \left(\mathbf{a}_{11}, \mathbf{a}_{12}, \ldots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \ldots, \mathbf{a}_{2K_2}, \ldots, \mathbf{a}_{IK_I}\right)$, define these linear combinations.

The multi-dimensional form of the model assumes that a set of D traits underlies the individuals' responses. The D latent traits define a D-dimensional latent space. The vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_D)'$, represents an individual's position in the D-dimensional latent space.

The model also introduces a scoring function that allows specifying the score or performance level assigned to each possible response category to each item. To do so, the notion of a response score b_{ijd} is introduced, which gives the performance level of an observed response in category j, item i, dimension d. The scores across D dimensions can be collected into a column vector $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ikD})^T$ and again collected into the scoring sub-matrix for item i, $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{iD})^T$ and then into a scoring matrix $\mathbf{B} = (\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T)^T$ for the entire test. (The score for a response in the zero category is zero, but, under certain scoring schemes, other responses may also be scored zero.) The scoring matrix, $\mathbf{B}$, represents the relationships between items and dimensions, and the design matrix, $\mathbf{A}$, represents the relationships between items and the model parameters.

The probability of a response in category j of item i is modelled as

9.2

$$\Pr\left(X_{ij} = 1; \mathbf{A}, \mathbf{B}, \xi \mid \boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{b}_{ij}\theta + \mathbf{a}_{ij}^T \xi\right)}{\sum_{k=1}^{K} \exp\left(\mathbf{b}_{ik}\theta + \mathbf{a}_{ik}^T \xi\right)}$$

There is a response vector,

9.3

$$f(\mathbf{x}, \xi \mid \theta) = \Psi(\theta, \xi) \exp[\mathbf{x}^T (\mathbf{B}\theta + \mathbf{A}\xi)]$$

with

9.4

$$\Psi(\boldsymbol{\theta},\boldsymbol{\xi}) = \left\{\sum_{\mathbf{z}\in\Omega}\exp\left[\mathbf{z}^T(\mathbf{B}\boldsymbol{\theta}+\mathbf{A}\boldsymbol{\xi})\right]\right\}^{-1}$$

where $\boldsymbol{\Omega}$ is the set of all possible response vectors.

The population model

The item response model is a conditional model, in the sense that it describes the process of generating item responses conditional on the latent variable, $\boldsymbol{\theta}$. The complete definition of the model, therefore, requires the specification of a density, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\alpha})$ for the latent variable, $\boldsymbol{\theta}$. Let $\boldsymbol{\alpha}$ symbolise a set of parameters that characterise the distribution of $\boldsymbol{\theta}$. The most common practice, when specifying uni-dimensional marginal item response models, is to assume that students have been sampled from a normal population with mean μ and variance σ^2.That is:

9.5

$$f_\theta(\boldsymbol{\theta};\boldsymbol{\alpha}) \equiv f_\theta(\theta;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right]$$

or equivalently

9.6

$$\theta = \mu + E$$

where $E \sim N(0,\sigma^2)$

Adams, Wilson and Wu (1997) discuss how a natural extension of [9.6] is to replace the mean, μ, with the regression model, $\mathbf{Y}_n^T\boldsymbol{\beta}$, where $\mathbf{Y}_n$ is a vector of u fixed and known values for student n, and β is the corresponding vector of regression coefficients. For example, $\mathbf{Y}_n$ could be constituted of student variables such as gender or socio-economic status. Then the population model for student n becomes

9.7

$$\theta_n = \mathbf{Y}_n^T\boldsymbol{\beta} + E_n$$

where it is assumed that the E_n are independently and identically normally distributed with mean zero and variance σ^2 so that [9.7] is equivalent to:

9.8

$$f_\theta(\theta_n;\mathbf{Y}_n,b,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(\theta_n-\mathbf{Y}_n^T\boldsymbol{\beta})^T(\theta_n-\mathbf{Y}_n^T\boldsymbol{\beta})\right]$$

a normal distribution with mean $\mathbf{Y}_n^T\boldsymbol{\beta}$ and variance σ^2. If [9.8] is used as the population model then the parameters to be estimated are $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\xi}$.

The generalisation needs to be taken one step further to apply it to the vector-valued $\boldsymbol{\theta}$ rather than the scalar-valued θ. The multi-dimensional extension results in the multivariate population model:

9.9

$$f_\theta(\boldsymbol{\theta}_n;\mathbf{w}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}_n-\boldsymbol{\gamma}\mathbf{w}_n)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_n-\boldsymbol{\gamma}\mathbf{w}_n)\right]$$

where $\boldsymbol{\gamma}$ is a $u \times D$ matrix of regression coefficients, $\boldsymbol{\Sigma}$ is a $D \times D$ variance-covariance matrix, and $\mathbf{W}_n$ is a $u \times 1$ vector of fixed variables.

In PISA, the $\mathbf{W}_n$ variables are referred to as conditioning variables.

Combined model

In [9.10], the conditional item response model [9.2] and the population model [9.9] are combined to obtain the unconditional, or marginal, item response model:

9.10

$$f_x\left(\mathbf{x};\boldsymbol{\xi},\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)=\int_{\boldsymbol{\theta}} f_x\left(\mathbf{x};\boldsymbol{\xi}\,|\,\boldsymbol{\theta}\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta};\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)d\boldsymbol{\theta}$$

It is important to recognise that under this model the locations of individuals on the latent variables are not estimated. The parameters of the model are $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\xi}$.

The procedures used to estimate model parameters are described in Adams, Wilson and Wu (1997), Adams, Wilson and Wang (1997), and Wu, Adams and Wilson (1997).

For each individual it is possible, however, to specify a posterior distribution for the latent variable, given by:

9.11

$$\begin{aligned}\mathrm{h}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n;\mathbf{W}_n,\boldsymbol{\xi},\boldsymbol{\gamma},\boldsymbol{\Sigma}\,|\,\mathbf{x}_n\right)&=\frac{f_x\left(\mathbf{x}_n;\boldsymbol{\xi}\,|\,\boldsymbol{\theta}_n\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n;\mathbf{W}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)}{f_x\left(x_n;\mathbf{W}_n,\boldsymbol{\xi},\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)}\\&=\frac{f_x\left(\mathbf{x}_n;\boldsymbol{\xi}\,|\,\boldsymbol{\theta}_n\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n;\mathbf{W}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)}{\int_{\boldsymbol{\theta}_n} f_x\left(\mathbf{x}_n;\boldsymbol{\xi}\,|\,\boldsymbol{\theta}_n\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n;\mathbf{W}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)}\end{aligned}$$

Plausible values

As with all item response scaling models, student proficiencies (or measures) are not observed; they are missing data that must be inferred from the observed item responses. There are several possible alternative approaches for making this inference. PISA uses the imputation methodology usually referred to as plausible values (PVs). PVs are a selection of likely proficiencies for students that attained each score. For each scale and subscale, five plausible values per student are included in the international database.

Using item parameters anchored at their estimated values from the international calibration, the plausible values are random draws from the marginal posterior of the latent distribution [9.11] for each student. For details on the uses of plausible values, see Mislevy (1991) and Mislevy et al. (1992).

In PISA, the random draws from the marginal posterior distribution are taken as follows.

Draw M vector-valued random deviates, $\left\{\boldsymbol{\varphi}_{mn}\right\}_{m=1}^{M}$, from the multivariate normal distribution, $f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n;\mathbf{W}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)$, for each case n, these vectors are used to approximate the integral in the denominator of [9.11], using the Monte-Carlo integration:[1]

9.12

$$\int_{\theta} f_x\left(\mathbf{x};\xi\,|\,\theta\right) f_{\theta}\left(\theta;\gamma,\Sigma,\mathbf{W}\right)d\theta\approx\frac{1}{M}\sum_{m=1}^{M} f_x\left(\mathbf{x};\xi,\varphi_{mn}\right)\equiv\Im$$

At the same time, the values

9.13

$$p_{mn}=f_x\left(x_n;\boldsymbol{\xi}\,|\,\boldsymbol{\varphi}_{mn}\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\varphi}_{mn};\mathbf{W}_n,\boldsymbol{\gamma},\boldsymbol{\Sigma}\right)$$

are calculated, so that we obtain the set of pairs $\left(\boldsymbol{\varphi}_{mn}, {}^{p_{mn}}\!\big/_{\Im}\right)_{m=1}^{M}$, which can be used as an approximation of the posterior density (11); and the probability that $\boldsymbol{\varphi}_{nj}$ could be drawn from this density is given by:

9.14

$$q_{nj} = \frac{p_{mn}}{\sum_{m=1}^{M} p_{mn}}$$

At this point, L uniformly distributed random numbers $\{\eta_i\}_{i=1}^{L}$ are generated, one for each required plausible vector; and for each random draw, the vector, $\boldsymbol{\varphi}_{ni_0}$, that satisfies the condition

9.15

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_i \le \sum_{s=1}^{i_0} q_{sn}$$

is selected as a plausible vector.

ANALYSIS OF DATA WITH PLAUSIBLE VALUES

It is very important to recognise that plausible values are not test scores and should not be treated as such. They are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual - that is, the marginal posterior distribution [9.11]. As such, plausible values contain random error variance components and are not as optimal as scores for individuals. Plausible values as a set are better suited to describing the performance of the population. This approach, developed by Mislevy and Sheehan (1987, 1989) and based on the imputation theory of Rubin (1987), produces consistent estimators of population parameters. Plausible values are intermediate values provided to obtain consistent estimates of population parameters using standard statistical analysis software such as *SPSS*® and *SAS*®. As an alternative, analyses can be completed using *ConQuest* (Wu, Adams and Wilson, 1997).

If an analysis with plausible values were to be carried out, then it would ideally be undertaken five times, once with each relevant plausible values variable. The results would be averaged, and then significance tests adjusting for variation between the five sets of results computed.

More formally, suppose that $r(\boldsymbol{\theta},\mathbf{Y})$ is a statistic that depends upon the latent variable and some other observed characteristic of each student. That is: $(\boldsymbol{\theta},\mathbf{Y}) = (\theta_1,y_1,\theta_2,y_2,\ldots,\theta_N,y_N)$ where (θ_n,y_n) are the values of the latent variable and the other observed characteristic for student n. Unfortunately, θ_{n} is not observed, although we do observe the item responses, x_n from which we can construct for each student n, the marginal posterior $h_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n; y_n, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma} \mid \mathbf{x}_n\right)$. If $h_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}; \mathbf{Y}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma} \mid \mathbf{X}\right)$ is the joint marginal posterior for $n = 1,\ldots N$ then we can compute:

9.16

$$\begin{aligned} r^*(\mathbf{X},\mathbf{Y}) &= E\left[r^*(\boldsymbol{\theta},\mathbf{Y}) \mid \mathbf{X},\mathbf{Y}\right] \\ &= \int_{\boldsymbol{\theta}} r(\boldsymbol{\theta},\mathbf{Y})\, h_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}; \mathbf{Y}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma} \mid \mathbf{X}\right) d\boldsymbol{\theta} \end{aligned}$$

The integral in [9.16] can be computed using the Monte-Carlo method. If M random vectors $(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_M)$ are drawn from $h_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}; \mathbf{Y}, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\Sigma} \mid \mathbf{X}\right)$ [9.16] is approximated by:

9.17

$$\begin{aligned} r^*(\mathbf{X},\mathbf{Y}) &\approx \frac{1}{M}\sum_{m=1}^{M} r(\boldsymbol{\theta}_m, \mathbf{Y}) \\ &= \frac{1}{M}\sum_{m=1}^{M} \hat{r}_m \end{aligned}$$

where $\hat{r}_m$ is the estimate of r computed using the m-th set of plausible values.

From [9.17] we can see that the final estimate of r is the average of the estimates computed using each randomly drawn vector in turn. If U_{m} is the sampling variance for $\hat{r}_m$ then the sampling variance of r^* is:

9.18

$$V = U^* + \left(1 + M^{-1}\right) B_M$$

where $U^* = \frac{1}{M}\sum_{m=1}^{M} U_m$ and $B_M = \frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{r}_m - r^*\right)^2$

An α-% confidence interval for r^* is $r^* \pm t_\upsilon\left((1-\alpha)/2\right)V^{1/2}$ where $t_\upsilon(s)$ is the s-percentile of the t-distribution with v degrees of freedom. $\upsilon = \left[\frac{f_M^2}{M-1} + \frac{\left(1-f_M\right)^2}{d}\right]^{-1}$, $f_M = \left(1 + M^{-1}\right)B_M/V$ and d is the degree of freedom that would have applied had θ_n been observed. In PISA, d will vary by country and have a maximum possible value of 80.

APPLICATION TO PISA

In PISA, the mixed coefficients multinomial logit model described above was used in three steps: national calibrations, international scaling and student score generation.

For both the national calibrations and the international scaling, the conditional item response model [9.2] is used in conjunction with the population model [9.9], but conditioning variables are not used. That is, it is assumed that students have been sampled from a multivariate normal distribution.

The design matrix was chosen so that the partial credit model (Masters, 1982) was used for items with multiple score categories and the simple logistic model was fitted to the dichotomously scored items.

National calibrations

National calibrations were performed separately, country by country, using unweighted data. Country means were constrained to zero during the estimation process. For the countries that administered booklet sets that included the core and standard items a linear transformation was applied to the national item difficulties so that the core and standard items have a mean of zero. For the countries that have used booklets that included core and easy items a linear transformation was applied to the national item difficulties so that the core items have the same mean as the mean of the core items for the OECD calibration sample. The results of these analyses, which were used to monitor the quality of the data and to make decisions regarding national item treatment, are given in Chapter 12.

The outcomes of the national calibrations were used to make a decision about how to treat each item in each country. This means that an item may be deleted from PISA altogether if it has poor psychometric characteristics in more than ten countries (a "dodgy" item); it may be deleted from the scaling in particular countries if it has poor psychometric characteristics in those particular countries but functions well in the vast majority of others. When reviewing the national calibrations, particular attention was paid to the fit of the items to the scaling model, item discrimination and item-by-country interactions.

Item response model fit (weighted mean square MNSQ)

For each item parameter, the *ConQuest* fit mean square index (Wu, 1997) was used to provide an indication of the compatibility of the model and the data. For each student, the model describes the probability of obtaining the different item scores. It is therefore possible to compare the model prediction and what has been observed for one item across students. Accumulating comparisons across students gives an item-fit statistic. As the fit statistics compare an observed value with a predicted value, the fit is an analysis of residuals. In the case of the item infit mean square, values near one are desirable. A weighted MNSQ greater than one is associated with a low discrimination index, meaning the data exhibits more variability than expected by the model, and an infit mean square less than one is associated with a high discrimination index, meaning the data exhibits less variability than expected by the model.

Discrimination coefficients

For each item, the correlation between the students' score and aggregate score on the set for the same domain and booklet as the item of interest was used as an index of discrimination. If p_{ij} (calculated as x_{ij}/m_i) is the proportion of score levels that student *i* achieved on item *j*, and $p_i = \sum_j p_{ij}$ (where the summation is of the items from the same booklet and domain as item *j*) is the sum of the proportions of the maximum score achieved by student *i*, then the discrimination is calculated as the product-moment correlation between p_{ij} and p_i for all students. For multiple-choice and short-answer items, this index will be the usual point-biserial index of discrimination.

The point-biserial index of discrimination for a particular category of an item is a comparison of the aggregate score between students selecting that category and all other students. If the category is the correct answer, the point-biserial index of discrimination should be higher than 0.20 (Ebel and Frisbie, 1986). They set out the following recommendations regarding the index of discrimination:

Magnitude	Comment	Recommended action for item
> 0.39	Excellent	Retain
0.30 – 0.39	Good	Possibilities for improvement
0.20 – 0.29	Mediocre	Need to check/review
0.00 – 0.20	Poor	Discard or review in depth
< -0.01	Worst	Definitely discard

Non-key categories should have a negative point-biserial index of discrimination. The point-biserial index of discrimination for a partial credit item should be ordered, i.e. categories scored 0 should have a lower point-biserial correlation than the categories scored 1, and so on.

Item-by-country interaction

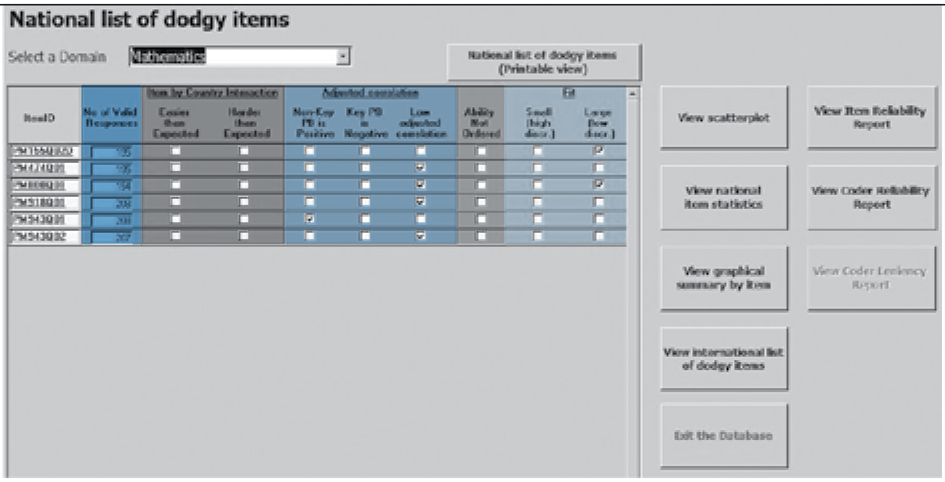
The national scaling provides nationally specific item parameter estimates. The consistency of item parameter estimates across countries was of particular interest. If the test measured the same latent trait per domain in all countries, then items should have the same relative difficulty or, more precisely, would fall within the interval defined by the standard error on the item parameter estimate (i.e. the confidence interval).

National reports

After national scaling was completed, all the available national item statistics were imported in the international item database. International level item statistics described next in this section were also included in this database. This allowed summarising national level statistics and performing the comparison to the international and aggregated item statistics. Database with national items statistics was returned to each participating country to assist in reviewing their data with the international contractor.

Figure 9.1 illustrates an interface of the national database. The main screen represents the interactive list of items by domain that are flagged as dodgy items in a country. Each column indicates a specific problem. Reliability and leniency reports will be discussed separately in Chapter 13.

■ Figure 9.1 ■
Main screen



Countries were asked to check the following statistics:

- Item by country interaction: The consistency of item parameters across countries is of particular importance in an international study. If the test measures the same underlying construct (or latent trait) the item should have similar relative difficulty in each country.
- Adjusted correlation: Adjusted correlation is a correlation between students' scores on an item and their adjusted domain scores, where the adjusted domain score is a student's total score for a domain minus the student's score for the particular item. For multiple-choice items this is equivalent to the point-biserial correlation of the correct response (key) and it should be 0.20 or higher. Otherwise it is marked as Low Adj. Correlation. If the item category is the key, the point biserial index should be positive (the same as for the item). Non-key categories (incorrect responses or distractors) should have a negative point biserial index.
- Ability not ordered: For partial credit items the student mean abilities should increase with increasing raw score; students that received score 0 should have lower mean abilities than those that had score 1 and those with score 2 should have higher mean abilities than those with 1.
- Fit: Infit Mean Square index is used to compare predicted value and observed value by analysis of residuals. Good fit should have values near one. An Infit Mean Square greater than one is associated with a low discrimination index while an Infit Mean Square lower than one is associated with a high discrimination index.

Four item reports could be generated using this database.

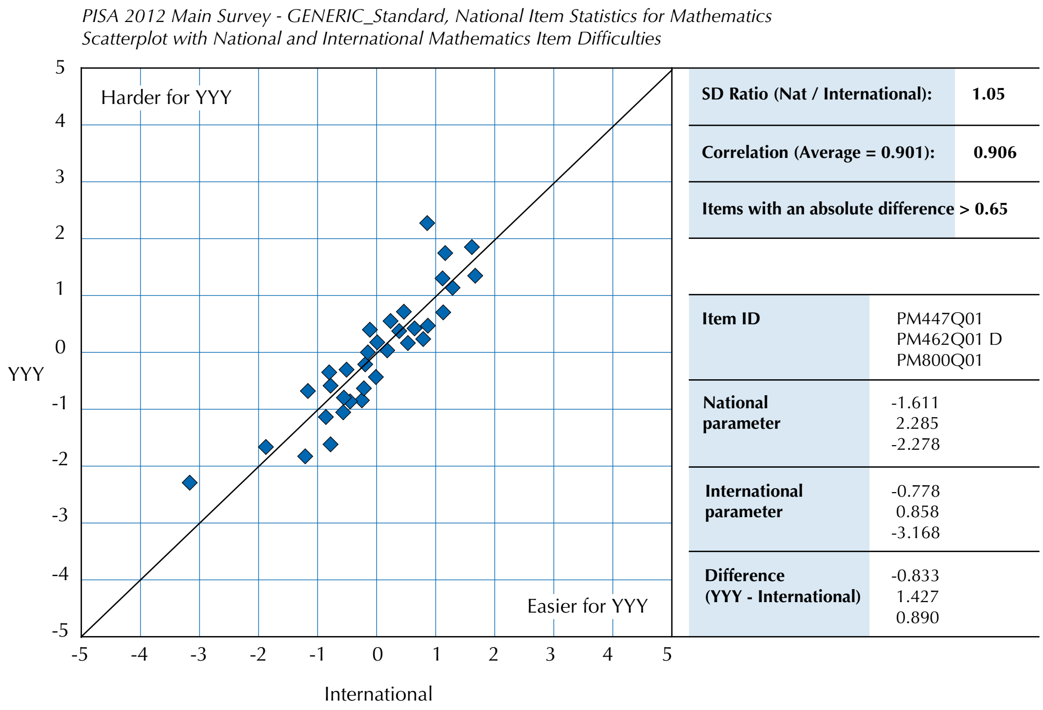
Report 1: Scatter plot

An example of a scatter plot report is given in Figure 9.2. This report shows the scatter plot of national and OECD/international item difficulties. Both sets of difficulties are centred on zero and are therefore referred to as relative difficulties. The vertical axis represents the national relative item difficulties and the horizontal axis the OECD or international relative item difficulties. Each dot is an item.

The scatter plot gives an overview of the behaviour of all items in a domain in one country compared to the pooled OECD set (500 students from each OECD country available at the time of analysis pooled together) or the international set (500 students from each country available at the time of analysis pooled together).

■ Figure 9.2 ■

Example of scatter plot



PISA 2012 Main Survey - GENERIC_Standard, National Item Statistics for Mathematics
Scatterplot with National and International Mathematics Item Difficulties

SD Ratio (Nat / International):	**1.05**
Correlation (Average = 0.901):	**0.906**
Items with an absolute difference > 0.65	

Item ID	PM447Q01 PM462Q01 D PM800Q01
National parameter	-1.611 2.285 -2.278
International parameter	-0.778 0.858 -3.168
Difference (YYY - International)	-0.833 1.427 0.890

Figure 9.2 provides an illustration of the overall level of agreement and it assists in identifying outliers. Items that lie exactly on the identity line (the diagonal line) have equal national and international relative item difficulties. An outlier occurs when the relative national item difficulty is very different from the OECD/international relative item difficulty. In Figure 9.2 there are a couple of obvious outliers. This suggests that something could be wrong with these items.

The table next to the scatter plot lists all items with an absolute difference of more than 0.65. The National Centres were asked to check these items carefully for any translation or printing errors.

There are two types of summary statistics displayed in the top right box of Figure 9.2:

- Standard deviation ratio compares the spread of national item difficulties to the spread of the OECD/international item difficulties. It should be close to 1.
- Correlation should be similar to the OECD average correlation.

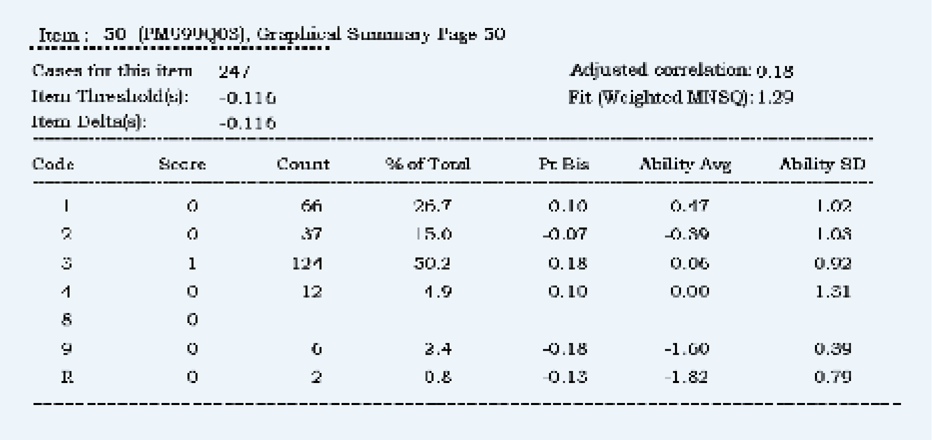
For this particular country both figures are satisfactory: the standard deviation ratio is sufficiently close to one and the correlation is sufficiently similar to the OECD average correlation.

Report 2: Descriptive statistics on individual items in tabular form

A detailed item-by-item report was provided in tabular form showing the basic item analysis statistics at the national level. This report provides classical item statistics for each item used in the national calibration. Summaries of item statistics are presented in a tabular form in item identifier order. If for any reason, an item is excluded from the national calibration, the item identifier will be listed at the end of the report. An example of item statistics for the fictitious item with identifier *PM999Q03* is shown in Figure 9.3.

■ Figure 9.3 ■

Example of item statistics in tabular form



```
Item :  50  (PM999Q03), Graphical Summary Page 50
------------------------------------------
Cases for this item      247                                Adjusted correlation: 0.18
Item Threshold(s):       -0.116                             Fit (Weighted MNSQ): 1.29
Item Delta(s):           -0.116
--------------------------------------------------------------------------------------------
Code         Score         Count       % of Total        Pt Bis      Ability Avg     Ability SD
--------------------------------------------------------------------------------------------
  1            0             66          26.7             0.10          0.47            1.02
  2            0             37          15.0            -0.07         -0.39            1.03
  3            1            124          50.2             0.18          0.06            0.92
  4            0             12           4.9             0.10          0.00            1.31
  8            0
  9            0              6           2.4            -0.18         -1.60            0.39
  R            0              2           0.8            -0.13         -1.82            0.79
-----------------------------------------------------------------------------------------------
```

Two hundred and forty seven students have responded to this item in this country.

The national threshold and delta (difficulty) are -0.116 (for dichotomous items these two values are always the same).

The item adjusted correlation is 0.18. This is lower than 0.2 and would be reported on the interactive list of dodgy items and in the graphical summary report that is described in the next section.

The weighted mean square (MNSQ) fit statistic is 1.29. Small variations around one are expected, however, values larger than 1.2 indicate that the item discrimination is lower than assumed by the model, and values below 0.8 show that the item discrimination is higher than assumed. In this particular case the item would have a tick on the interactive screen in the 'Large Fit' column and in the graphical summary report that is described in the next section.

The first column gives the original responses. This is a multiple-choice item and therefore, the responses are: 1=A, 2=B, 3=C, 4=D, 8='invalid', 9='missing' and R='not reached'. Please note that there are no statistics for code 8. This is because there were no students in this country who gave invalid responses to this item.

The second column shows the score assigned to each response category. The correct response to this item is 3 (C).

The third and fourth columns in the table list the number and percentage of students in each category. In this country, 124 students (50.2%) gave the correct response.

The point-biserial correlations are presented in column five. This is the correlation between a response category coded as a dummy variable (a score of 1 for students that responded with the current code and a score of 0 for students in other response categories) and the total domain score. For dichotomous items the point-biserial is equal to the adjusted correlation (0.18 in Figure 9.3). Correct responses should have positive correlations with the total score, incorrect responses negative correlations. In this case one of the incorrect responses (4) has positive point-biserial (0.10). However the item would not have a tick on the interactive screen in the corresponding column for positive point biserial in non-key category, because there were fewer than 15 students who responded to distractor 4. Rather, this item would be flagged for low adjusted correlation less than 0.20.

The two last columns show the average ability of students responding in each category and the associated standard deviation. The average ability is calculated by domain. If an item is functioning well the group of students that gave the correct response should have a higher mean ability than the groups of students that provided each of the incorrect responses. This is true for categories 1 and 2. For category 4 this doesn't hold, but since the number of students is less than fifteen, this is not flagged.

Report 3: Graphical summary of descriptive statistics by item

This report provides comparisons between national and international item statistics in graphical form, one page per item.

An example of a full page for one item is given in Figure 9.4. More detailed information about each part of this report labelled A to D follows.

Part A

The top table in Figure 9.4 starts with the item code followed by the item name and item number (*CM999Q01*: Graph Example Q1).[2] For mathematics items, there is also a group identifier on the right hand side of the table. In the PISA 2012 Main Survey, the majority of mathematics items (common and link items) were administered in all participating countries. Seventeen countries used booklets that included a set of easier items. This was done to better cover the range of abilities in every country.

Item identifiers are followed by the overall item statistics, the same as in the national item statistics report described in the previous section: number of cases, adjusted correlation, weighted (infit) mean square (MNSQ), item thresholds and item difficulty (delta). In addition, item type (e.g. multiple choice) is presented. For multiple-choice items a key (correct choice) is also shown. Graph Example Q1 in Figure 9.4 is a partial credit item and therefore the key is not shown. Processing the responses to items of this type usually required manual coding.

The next section of part A contains national, international and OECD statistics by response category. The first row contains the score for each category, the second and third rows contain number of students and percentage of students in each category in the country. OECD % is the percentage of students in each category in the pooled OECD data. INT % is the percentage of students in the category in the pooled data of all countries that administered the item. Note that OECD % is not available for the set of easier mathematics items and financial literacy items.

Average ability (ability avg), standard deviation ability (ability SD), and point-biserial (pt bis) are the same national statistics as in the national item statistics report. These statistics were described in the previous section.

Part B

The displayed graphs in part B facilitate the process for identifying the possible national anomalies related to item statistics by response category.

The first graph is important for partial credit items. It helps to check whether the average ability increases with the score points, as shown in Figure 9.4. Note that categories "9" and "R" are not identified as score points.

The second graph is important for multiple-choice items. It helps to check whether:

- a non-key category has a positive point-biserial;
- a non-key category has a point-biserial higher than the key category; or
- the key category has a negative point-biserial.

■ Figure 9.4 ■

Example of graphical summary by item

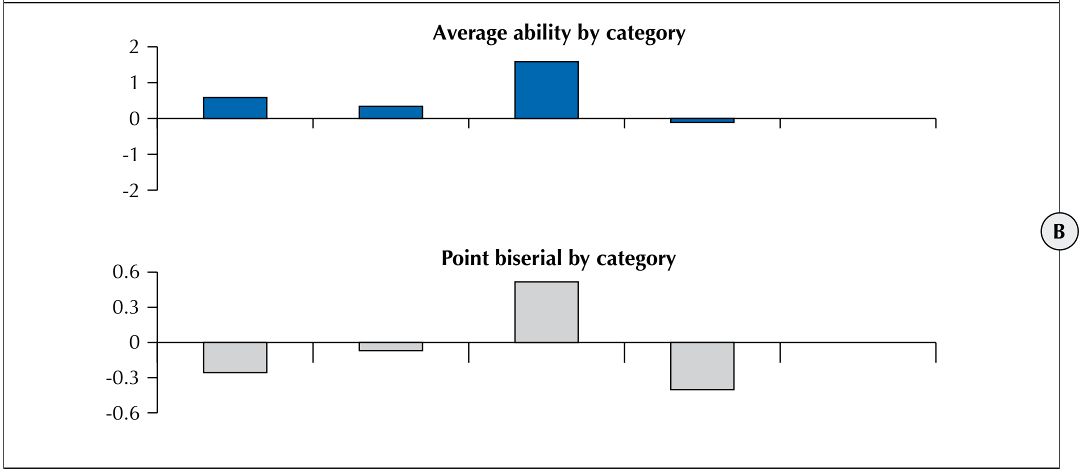
PISA MS12: Graphical presentation of item statistics for Country

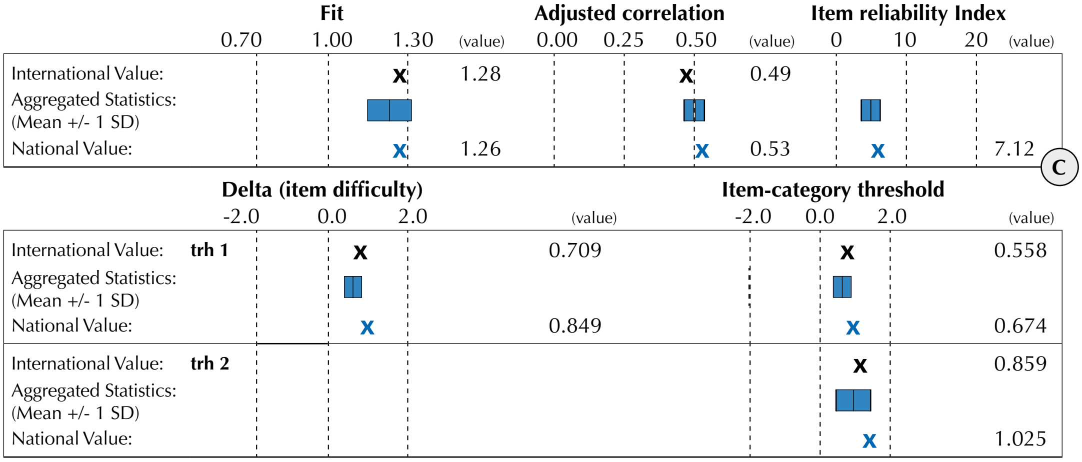
CM999Q01: Graph Example Q1 (Item Format: Constructed Reponse Expert)

Number of Cases: 235	*Adjusted Correlation: 0.53*	*Item Threshold(s): 0.674*	*1.025*
Item Type: Partial Credit Item	*Fit (Weighted MNSQ): 1.26*	*Item Delta(s): 1.891*	*-0.192*

Response	**0**	**1**	**2**	**9**	**R**
Score	0	1	2	0	0
Students	70	26	115	24	
Percentage of students	29.79	11.06	48.94	10.21	
OECD %	30.42	9.91	50.06	9.54	0.06
INT %	30.73	12.45	45.12	10.59	0.12
Ability Avg	0.56	0.34	1.55	-0.1	
Ability SD	0.86	0.94	0.89	0.94	
Pt Bis	-0.27	-0.06	0.52	-0.39	

A





D

	Item by country interaction		**Adjusted correlation**				**Fit**	
No of Countries	**Easier than Expected**	**Harder than Expected**	**Non-key PB is Positive**	**Key PB is Negative**	**Low Adjusted Correlation**	**Ability not Ordered**	**Small (High Discrimination Item)**	**Large (Low Discrimination Item)**
CM999Q01	☐	☐	☐	☐	☐	☐	☐	☑
Countries: 48	12	9	0	0	0	2	0	23
OECD countries: 22	4	7	0	0	0	2	0	15
Other countries: 26	8	2	0	0	0	0	0	8

Part C

This part presents the graphical comparisons of overall item statistics at the national and OECD level.

National scaling provides for each country and item, the weighted MNSQ, adjusted correlation, delta item parameter estimate (or difficulty estimate) and threshold estimates. Item Reliability Index would also be provided if the item is a constructed response item that requires multiple coding as part of the process of evaluating the reliability of items and coders. For each item these national values will be compared with the pooled OECD value and average value for all OECD countries in the database at the time of comparison.

The black crosses at the top of each of the pictures represent the value of the coefficients computed from the pooled OECD data. The blue rectangles show the distribution of values obtained from each of available OECD country (all students). To obtain this distribution each OECD country is calibrated separately. Then the mean and standard deviation of the national estimates are computed. The rectangles are located so that their mid-point (indicated with a vertical bar) is at the mean and the left and right boundaries are located at the mean plus and minus one standard deviation respectively.

The blue crosses at the bottom of the pictures indicate the values computed only for the national dataset.

Any substantial differences between the national value and the OECD value, or the average OECD value, indicate that the item is behaving differently in that country in comparison to the other countries. This might reflect a mistranslation or printing problem. On the other hand, if the item is misbehaving in many countries, it might reflect a specific problem in the source item and not with one or more national versions of this item.

OECD statistics are not available for the easier mathematics items. Hence, the statistics for these items are calculated based on the pooled data from 17 countries.

Part D

At the bottom of the page a table with check boxes shows whether any substantial problems were found as a result of the national calibration for the particular item. The table indicates if an item was flagged for one of the following reasons:

- the relative national item difficulty is significantly higher or lower than OECD/International relative item difficulties;
- for multiple-choice items one of the non-key categories has a point-biserial correlation higher than 0.05 (only reported if the category was chosen by at least 15 students);
- for multiple-choice items the key category has a point-biserial lower than –0.05 (only reported if the category was chosen by at least 15 students);
- the adjusted correlation of the item is lower than 0.2;
- for partial credit items the category abilities are not ordered (only reported if both score categories in comparison have at least 15 students each); or
- the fit statistics are higher than 1.2 or lower than 0.8.

In the example in Figure 9.4, the box is ticked indicating large fit index. This is also shown in Part A (weighted MNSQ=1.26).

The next row below the tick boxes shows how many countries in total have a similar problem for the same item. The last two rows are the numbers of OECD countries and partner countries that have the same problem. The large fit problem, which is identified in Parts A and D, does not look problematic on the graph in Part C for this particular country. It is because out of 48 available countries, 23 countries (or 15 out of 22 available OECD countries) have the same problem (the figures are fictitious). This indicates a specific problem in the source item instead of possible mistranslation or misprint problems in the national versions.

However, if an item has at least one tick, and the number of countries below this tick is less than 10, the National Centres were strongly recommended to review the translation and printing of the item in all booklets and its appropriateness for the national context.

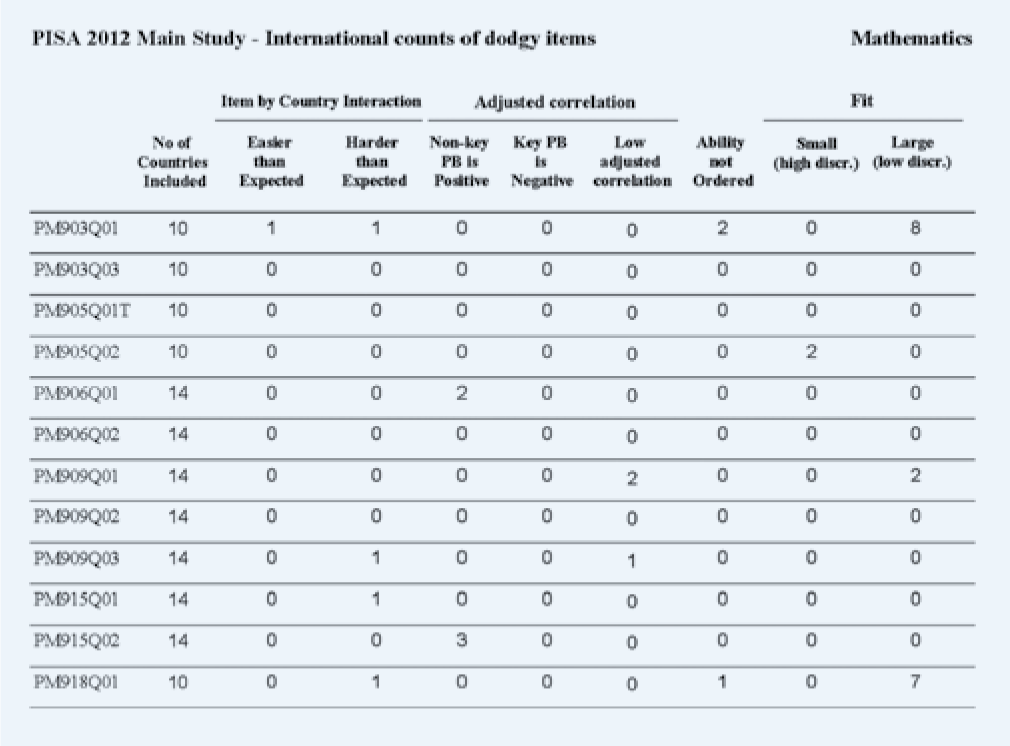
All flagged items are considered to be dodgy items either nationally if a problem occurs only in a particular country, or internationally if the same problem occurs in many countries (in more than 50% of cases).

Report 4: International list of dodgy items

The last report gives a summary of dodgy items for all countries included in the analysis at the time of reporting. A part of this table is given in Figure 9.5, showing items for which no indication of problems is evident, and items for which some such indication is present.

■ Figure 9.5 ■

Example of an international list of dodgy items



PISA 2012 Main Study - International counts of dodgy items — Mathematics

		Item by Country Interaction		Adjusted correlation				Fit	
	No of Countries Included	Easier than Expected	Harder than Expected	Non-key PB is Positive	Key PB is Negative	Low adjusted correlation	Ability not Ordered	Small (high discr.)	Large (low discr.)
PM903Q01	10	1	1	0	0	0	2	0	8
PM903Q03	10	0	0	0	0	0	0	0	0
PM905Q01T	10	0	0	0	0	0	0	0	0
PM905Q02	10	0	0	0	0	0	0	2	0
PM906Q01	14	0	0	2	0	0	0	0	0
PM906Q02	14	0	0	0	0	0	0	0	0
PM909Q01	14	0	0	0	0	2	0	0	2
PM909Q02	14	0	0	0	0	0	0	0	0
PM909Q03	14	0	1	0	0	1	0	0	0
PM915Q01	14	0	1	0	0	0	0	0	0
PM915Q02	14	0	0	3	0	0	0	0	0
PM918Q01	10	0	1	0	0	0	1	0	7

If an item has poor psychometric properties in a large number of countries then it most likely should be explained by reasons other than mistranslation and misprint.

International calibration

There were new elements added to the PISA 2012 paper-based assessment of mathematics, reading and science; computer-based assessment of problem solving; digital reading and computer-based mathematics; financial literacy and reading components.

Since PISA 2000 international item parameters were set by applying the conditional item response model [9.2] in conjunction with the multivariate population model [9.9], without using conditioning variables, to a sub-sample of students. Traditionally a subsample of students referred to as an OECD calibration sample was formed using 500 students drawn at random from each of the OECD participating countries. In PISA 2009 countries with an expected mean reading score less than 450 were given the option to choose an easier set of booklets for the Main Survey. In total, 20 countries opted for the easier booklets, of which two, Mexico and Chile, were OECD member countries. This required creation of an extra calibration sample that included non-OECD countries.

In 2009, reading items required a two-step calibration process.

Step 1: The core and standard items were calibrated using the OECD calibration sample, which contained 500 students from 34 OECD countries.

Step 2: The easier items that were not included in the regular booklets were calibrated using the easy booklets calibration sample, while anchoring the core and standard items to the estimates obtained from Step 1. The easy booklets calibration sample was formed by adding subsamples of 500 students from each of the 20 countries that used the easy booklets to the international OECD calibration sample.

In PISA 2012, the use of data for calibration purposes from OECD countries only was not viable because not a sufficient number of the same OECD countries were participating in all options. Calibration of all optional parts was based on all countries that participated in those options. Consistent with the treatment of options, it was decided that the calibration is based on data from all available countries for the PISA 2012 paper-based assessment.

Similarly to PISA 2009, in PISA 2012 countries with an expected mean mathematics score less than 450 were given the option to choose an easier set of booklets for the Main Survey (see Chapter 2 for more details). In total, 17 countries opted for the easier booklets, of which two, Mexico and Chile, were OECD member countries. This subsample of students, referred to as an international calibration sample, consisted of 31 500 students comprising 500 students drawn at random from each of the 63 participating countries.[3] Not-reached items were excluded from the calibration. For model identification the average difficulty of all items in each domain was set to zero.

For the options it was decided to create a calibration sample with a similar number of responses per item as for the paper-based test. For the paper-based test sampling 500 students yields 154 responses per item, since each student responds to approximately 4/13 of all items. This resulted in the following sample sizes for each PISA 2012 option:

- Problem Solving (PS)
  - 308 (=154×8/4) students per country that implemented PS only
  - 924 (=154×24/4) students per country that implemented PS and CBA (mathematics and reading)
- Computer-Based Assessment (CBA)
  - 924 (=154×24/4) students per country for mathematics
  - 462 (=154×24/8) students per country for reading
- Financial Literacy
  - 154 (=154× 4/4) students per country

The allocation of each PISA item to one of the PISA 2012 scales and corresponding item parameters are given in Annex A.

Student score generation

Five multi-dimensional scaling models were used in the PISA 2012 Main Survey. The first model, made up of one reading, one science and one mathematics dimension, was used for reporting overall scores for reading, science and mathematics. A second model, made up of one science, one reading and four mathematics scales, was used to generate scores for the four mathematics subscales *Change and Relationships, Quantity, Space and Shape, Uncertainty and Data*. A third model, made up of one science, one reading and three mathematics scales was used to generate scores for the three mathematics subscales: *Employ, Formulate* and *Interpret*. A fourth model, made up of one reading, one science, one mathematics, one digital reading dimension, one digital mathematics and one digital problem solving dimension was used for reporting overall scores for reading, science, mathematics and computer-based mathematics, digital reading and computer problem solving scales for countries that implemented the computer-based assessment (CBA) in the PISA 2012 Main Survey. A fifth model, made up of one reading, one science, one mathematics and one digital problem solving dimension was used for reporting overall scores for reading, science, mathematics and computer problem solving scales for those countries that implemented problem solving in the PISA 2012 Main Survey as the only computer-based component.

The first three models were implemented in one step and the last two models, for countries that participated in CBA, were implemented in two steps, as it will be described later.

Sixty-five plausible values, five for each of the 13 PISA 2012 scales are included in the PISA 2012 database. *PV1MATH* to *PV5MATH* are for mathematical literacy; *PV1SCIE* to *PV5SCIE* for scientific literacy, *PV1READ* to *PV5READ* for reading literacy, *PV1CPRO* to *PV5CPRO* for computer problem solving assessment, *PV1CMAT* to *PV5CMAT* for the computer-based mathematics assessment and *PV1CREA* to *PV5CREA* for digital reading assessment. For the four mathematics content subscales, *change and relationships, quantity, space and shape, uncertainty and data,* the plausible values variables are *PV1MACC* to *PV5MACC, PV1MACQ* to *PV5MACQ, PV1MACS* to *PV5MACS, and PV1MACU* to *PV5MACU* respectively. For the three mathematics process subscales e*mploy, formulate and interpret*, the plausible values variables are *PV1MAPE* to *PV5MAPE, PV1MAPF* to *PV5MAPF, and PV1MAPI* to *PV5MAPI* respectively.

Constructing conditioning variables

The PISA conditioning variables are prepared using procedures based on those used in the United States National Assessment of Educational Progress (Beaton, 1987) and in IEA Third International Mathematics and Science Study

(Macaskill, Adams and Wu, 1998). All available student-level information, other than their responses to the items in the booklets, is used either as direct or indirect regressors in the conditioning model. The preparation of the variables for the conditioning proceeds as follows.

Variables for booklet identifier were represented by deviation contrast codes and were used as direct regressors. Each booklet was represented by one variable, except for reference booklet 13. Booklet 13 was chosen as reference booklet because it included items from all domains. The difference between simple contrast codes that were used in PISA 2000 and PISA 2003 is that with deviation contrast coding the sum of each column is zero (except for the UH – one hour – booklet), whereas for simple contrast coding the sum is one. The contrast coding scheme is given in Annex B. In addition to the deviation contrast codes, regression coefficients between reading or science and the booklet contrasts that represent booklets without science or reading were fixed to zero. The combination of deviation contrast codes and fixing coefficients to zero resulted in an intercept in the conditioning model that is the grand mean of all students that responded to items in a domain if only the booklet is used as independent variable. This way, the imputation of abilities for students that did not respond to any science or reading items is based on information from all booklets that have items in a domain and not only from the reference booklet as in simple contrast coding.

Other direct variables in the regression are gender (and missing gender if there are any) and deviation contrast codes for schools with the largest school as reference school, grade, mother and father *ISEI* (International Socio-Economic Index). All other categorical variables from the student, ICT (information and communication technology), ECQ (educational career questionnaire) and parent questionnaire were dummy coded. These dummy variables and all numeric variables (the questionnaire indices) were analysed in a principal component analysis. The details of recoding the variables before the principal component analysis are listed in Annex B. The number of component scores that were extracted and used in the scaling model as indirect regressors was country specific and explained 95% of the total variance in all the original variables.

The item-response model was fitted to each national data set and the national population parameters were estimated using item parameters anchored at their international location, the direct and indirect conditioning variables described above and fixed regression coefficients between booklet codes and the minor domains that were not included in the corresponding booklet.

For the countries with very large samples over 10 000 students the sample was divided into smaller data sets using either stratification variables or any other national variable that allow clearly identify distinct groups for example test of the language variable was used in some countries.

Given that the CBA reporting scale cannot influence the PISA paper-based assessment, it was suggested that the plausible values for computer-based assessment countries are drawn in two steps. The first model is a three-dimensional model with reading, mathematics and science. This model was used to estimate the regression coefficients for the background variables for three main domains. Subsequently final plausible values for all domains have been drawn from a four or six dimensional models including computer-based assessment and anchoring regression coefficients to the parameters from the three-dimensional paper-based model.

All students from schools that are sampled for computer-based assessment received plausible values for paper-based PISA and plausible values for computer-based assessment.

Five multi-dimensional scaling models described above were estimated.

BOOKLET EFFECTS

As with PISA 2003, PISA 2006 and PISA 2009, the PISA 2012 test design was balanced, so that the item parameter estimates that are obtained from scaling are not influenced by a booklet effect, as was the case in PISA 2000. However, due to the different location of domains within each of the booklets it was expected that there would still be booklet influences on the estimated proficiency distributions.

Modelling the order effect in terms of item positions in a booklet or at least in terms of cluster positions in a booklet would result in a very complex model. For the sake of simplicity in the international scaling, the effect was modelled separately for each domain at the booklet level, as in previous cycles.

To correct the student mathematics, reading and science scores for the booklet effects, two alternatives were considered:

- correcting all students' scores using one set of the internationally estimated booklet parameters; or
- correcting the students' scores using nationally estimated booklet parameters for each country.

When choosing between these two alternatives a number of issues were considered. First, it is important to recognise that the sum of the booklet correction values is zero for each domain, so the application of either of the above corrections does not change the country means or rankings. Second, if a national correction was applied then the booklet means will be the same for each domain within countries. As such, this approach would incorrectly remove a component of expected sampling and measurement error variation. Third, the booklet corrections are essentially an additional set of item parameters that capture the effect of the item locations in the booklets. In PISA all item parameters are treated as international values so that all countries are therefore treated in exactly the same way. Perhaps the following scenario best illustrates the justification for this. Suppose students in a particular country found the reading items on a particular booklet surprisingly difficult, even though those items have been deemed as central to the PISA definition of PISA literacy and have no technical flaws, such as a translation or coding error. If a national correction were used then an adjustment would be made to compensate for the greater difficulty of these items in that particular country. The outcome would be that two students from different countries who responded in the same way to these items would be given different proficiency estimates. This differential treatment of students based upon their country has not been deemed as suitable in PISA. Moreover this form of adjustment would have the effect of masking real underlying differences in literacy between students in those two countries, as indicated by those items.

Applying an international correction was therefore deemed the most desirable option from the perspective of cross-national consistency.

When estimating the item parameters, booklet effects were included in the measurement model to prevent confounding item difficulties and booklet effects. For the *ConQuest* model statement, the calibration model was:

item + item*step + booklet.

The booklet parameter, formally defined in the same way as item parameters, reflects booklet difficulty. This calibration model was used to estimate the international item parameters for mathematics, reading and science.

The booklet parameters obtained from this analysis were not used to correct for the booklet effect. Instead, a set of booklet parameters for the standard booklets was obtained by scaling the entire data set of equally weighted countries using booklet as a conditioning variable. The students who responded to the UH booklet were excluded from the estimation. A set booklet parameter for the easy booklets was obtained by scaling the entire set of equally weighted countries that opted to use an easy booklet set, using booklet as a conditioning variable.

The booklet parameter estimates obtained are reported in Chapter 12. The booklet effects are the amount that must be added to or subtracted from the proficiencies of students who responded to each booklet.

As the computer-based assessment test was balanced and only included two clusters of 20 minutes it was found that it was unnecessary to add a set of booklet parameters to the model and estimate a booklet effect.

DEVELOPING COMMON SCALES FOR THE PURPOSES OF TRENDS

The reporting scales that were developed for each of the domains of reading, mathematics and science in PISA 2000 were linear transformations of the natural logit metrics that result from the scaling as described above. The transformations were chosen so that the mean and standard deviation of the PISA 2000 scores was 500 and 100 respectively, for the equally weighted 27 OECD countries that participated in PISA 2000 that had acceptable response rates (Wu and Adams, 2002).

For PISA 2003, the decision was made to report the reading and science scores on these previously developed scales. That is, the reading and science reporting scales used for PISA 2000 and PISA 2003 are directly comparable. The value of 500, for example, has the same meaning as it did in PISA 2000.

For mathematics this was not the case, however. Mathematics, as the major domain, was the subject of major development work for PISA 2003, and the PISA 2003 mathematics assessment was much more comprehensive than the PISA 2000 mathematics assessment – the PISA 2000 assessment covered just two (*space and shape*, and *change and relationships*) of

the four areas that are covered in PISA 2003. Because of this broadening in the assessment it was deemed inappropriate to report the PISA 2003 mathematics scores on the same scale as the PISA 2000 mathematics scores. For mathematics the linear transformation of the logit metric was chosen such that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2003. For PISA 2006 the decision was made to report the reading on these previously developed scales. That is the reading reporting scales used for PISA 2000, PISA 2003 and PISA 2006 are directly comparable. Mathematics reporting scales are directly comparable for PISA 2003 and PISA 2006. For science a new scale was established in 2006. The metric for that scale was set so that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2006.

To permit a comparison of the PISA 2006 science results with the science results in previous data collections a science link scale was prepared. The science link scale provides results for 2003 and 2006 using only those items that were common to the two PISA studies. These results are provided in a separate database.

For PISA 2009, the decision was made to report the reading, mathematics and science scores on these previously developed scales. That is the reading scales used for PISA 2000, PISA 2003. PISA 2006 and PISA 2009 are directly comparable. PISA 2009 mathematics reporting scale is directly comparable to PISA 2003 and PISA 2006 and the science reporting scale is directly comparable to PISA 2006 scale.

Again for PISA 2012 the decision was made to report the reading, mathematics and science scores on these previously developed scales. That is the reading scales used for PISA 2000, PISA 2003. PISA 2006, PISA 2009 and PISA 2012 are directly comparable. PISA 2012 mathematics reporting scale is directly comparable to PISA 2003, PISA 2006 and PISA 2009 and the science reporting scale is directly comparable to PISA 2006 and PISA 2009 scale.

Further details on the various PISA reporting scales are given in Chapter 12.

Linking PISA 2012 for mathematics, reading, science and digital reading

The linking of PISA 2012 mathematics, reading and science to the existing scales was undertaken using standard common item equating methods.

The steps involved in linking the PISA 2009 and PISA 2012 scales were as follows:

Step 1: Item parameter estimates were obtained from the PISA 2012 calibration sample.

Step 2: A shift constant was computed to place the above item parameters estimates on the PISA 2009 scale so that the mean of the item parameter estimates for the common items was the same in 2012 as it was in 2009.

Step 3: The 2012 student abilities were estimated with item parameters anchored at their 2012 values.

Step 4: The above estimated student abilities were transformed with the shift computed in Step 2.

Note that this is a much simpler procedure than that which was employed in linking the reading and science between PISA 2003 and PISA 2000. The simpler procedure could be used on this occasion because the test design was balanced since PISA 2003 onwards.

Uncertainty in the link

In each case the transformation that equates the 2012 data with previous data depends upon the change in difficulty of each of the individual link items and as a consequence the sample of link items that have been chosen will influence the choice of transformation. This means that if an alternative set of link items had been chosen the resulting transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items, just as there is an uncertainty in values such as country means due to the use of a sample of students.

The uncertainty that results from the link-item sampling is referred to as linking error and this error must be taken into account when making certain comparisons between the results from different PISA data collection. Just as with the error that is introduced through the process of sampling students, the exact magnitude of this linking error cannot be determined. We can, however, estimate the likely range of magnitudes for this error and take this error into account when interpreting PISA results. As with sampling errors, the likely range of magnitude for the errors is represented as a standard error.

In PISA 2003 the link error was estimated as follows.

Let $\hat{\delta}_i^{2000}$ be the estimated difficulty of link i in PISA 2000 and let $\hat{\delta}_i^{2003}$ be the estimated difficulty of link i in PISA 2003, where the mean of the two sets of difficulty estimates for all of the link items for a domain is set at zero. We now define the value:

9.19

$$c_i = \hat{\delta}_i^{2003} - \hat{\delta}_i^{2000}$$

The value c_i is the amount by which item i deviates from the average of all link items in terms of the transformation that is required to align the two scales. If the link items are assumed to be a random sample of all possible link items and each of the items is counted equally then the link error can be estimated as follows:

9.20

$$error_{2000,2003} = \sqrt{\frac{1}{L}\sum c_i^2}$$

Where the summation is over the link items for the domain and L is the number of link items.

Monseur and Berezner (2007) have shown that this approach to the link error estimation is inadequate in two regards. First, it ignores the fact that the items are sampled as units and therefore a cluster sample rather than a simple random sample of items should be assumed. Secondly, it ignores the fact that partial credit items have a greater influence on students' scores than dichotomously scored items. As such, items should be weighted by their maximum possible score when estimating the equating error.

To improve the estimation of the link error the following improved approach has been used in PISA 2006. Suppose we have L link items in K units. Use i to index items in a unit and j to index units so that $\hat{\delta}_{ij}^{y}$ is the estimated difficulty of item i in unit j for year y, and let

9.21

$$c_{ij} = \hat{\delta}_{ij}^{2006} - \hat{\delta}_{ij}^{2003}$$

The size (total number of score points) of unit j is m_j so that:

$$\sum_{j=1}^{K} m_j = L \quad \text{and} \quad \bar{m} = \frac{1}{K}\sum_{j=1}^{K} m_j$$

Further let:

9.22

$$c_{\cdot j} = \frac{1}{m_i}\sum_{i=1}^{m_j} c_{ij} \quad \text{and} \quad \bar{c} = \frac{1}{K}\sum_{j=1}^{K} c_{\cdot j}$$

and then the link error, taking into account the clustering is as follows:

9.23

$$error_{2006,2003} = \sqrt{\frac{\sum_{j=1}^{K} m_j^2 \left(c_{\cdot j} - \bar{c}\right)^2}{K(K-1)\bar{m}^2}}$$

The PISA 2006 approach for estimating the link errors was used again in PISA 2009.

The most obvious example of a situation where there is a need to use linking error is in the comparison of the mean performance for a country between two PISA data collections. For example, let us consider a comparison between 2003 and 2009 of the performance of Norway in mathematics. The mean performance of Norway in 2003 was 495 with a standard error of 2.38, while in 2009 the mean was 498 with a standard error of 2.40.

The standard error on this difference, as mentioned above, is influenced by the linking error. The standard error is therefore equal to:

9.24

$$SE = \sqrt{\sigma^2_{\hat{\mu}2003} + \sigma^2_{\hat{\mu}2009} + error^2_{2003,2009}}$$

$$SE = \sqrt{2.38^2 + 2.40^2 + 1.99^2} = 3.92$$

The standardised difference in the Norwegian mean is 0.71, which is computed as follows:

$$0.71 = \frac{498 - 495}{3.92}$$

and is not statistically significant (absolute values less than 1.96 are not statistically significant at the 95% level of confidence).

In PISA 2012 the same method was used for link errors involving only 2 survey administrations. If however we are considering trends over more than 2 survey administrations, we should consider in addition the covariance between link errors. For example, consider now the correlation of linking errors for 2006 and 2009 referred to 2003. Let $c_{ij} = \hat{\delta}_{ij}^{2006} - \hat{\delta}_{ij}^{2003}$ and $d_{ij} = \hat{\delta}_{ij}^{2009} - \hat{\delta}_{ij}^{2003}$ and suppose we have L scores points for link items common to all 3 cycles in K units, then

9.25

$$\text{cov\_}\,2003_{2006,2009} = \frac{\sum_{j=1}^{K} m_j^2 \left(c_{.j} - \bar{c}\right)\left(d_{.j} - \bar{d}\right)}{K\left(K-1\right)\bar{m}^2}$$

Suppose we are looking at the comparison of mean performance for a country between several PISA survey administrations, say for mathematics from 2003 to 2009. Consider the PISA 2006 and 2009 administrations referred back to PISA 2003. The standard error of the difference between 2006 and 2003 will be

9.26

$$SE(diff_{2003,2006}) = \sqrt{error^2_{2003,2006} + \sigma^2_{2003} + \sigma^2_{2006}}$$

where $error_{2003,2006}$ is the linking error from above and σ_{2003} is the standard error of the mean. An analogous result is obtained for the differences between 2009 and 2003. The covariance between any of these two differences will involve the covariance between the link errors as given above and the standard error of 2003, so the covariance between the differences 2006 versus 2003 and 2009 versus 2003 will be

9.27

$$Cov(diff_{2003,2006}, diff_{2003,2009}) = cov\_\,2003_{2006,2009} + \sigma^2_{2003}$$

with similar results for the other two covariances between the differences. Suppose we wished to test if the sizes of two consecutive trends say from 2003 to 2006, and 2006 to 2009, were significantly different. We use:

9.28

$$\delta_{2009} - \delta_{2006} = (\delta_{2009} - \delta_{2003}) - (\delta_{2006} - \delta_{2003})$$

so that the link errors are referred back to 2003:

9.29

$$
\begin{aligned}
Cov(diff_{2003,2006}, diff_{2006,2009}) &= E((\varepsilon_{2006} - \varepsilon_{2003} + \delta_{2006} - \delta_{2003})(\varepsilon_{2009} - \varepsilon_{2006} + \delta_{2009} - \delta_{2006})) \\
&= E((\varepsilon_{2006} - \varepsilon_{2003})(\varepsilon_{2009} - \varepsilon_{2006})) + E((\delta_{2006} - \delta_{2003})(\delta_{2009} - \delta_{2003})) - E((\delta_{2006} - \delta_{2003})(\delta_{2006} - \delta_{2003})) \\
&= -\sigma^2_{2006} + Cov\_2003_{2006,2009} - error^2_{2003,2006}
\end{aligned}
$$

Now the variance and standard error of the difference between the 2 trends will be

9.30

$$
\begin{aligned}
Var(diff_{2006,2009} - diff_{2003,2006}) &= Var(diff_{2006,2009})^2 + Var(diff_{2003,2006})^2 - 2Cov(diff_{2006,2009}, diff_{2003,2006}) \\
&= \sigma^2_{2006} + \sigma^2_{2009} + error^2_{2006,2009} + \sigma^2_{2003} + \sigma^2_{2006} + error^2_{2003,2006} \\
&\quad - 2(-\sigma^2_{2006} - error^2_{2003,2006} + \mathrm{cov}\_2003_{2006,2009}) \\
&= 4\sigma^2_{2006} + \sigma^2_{2009} + \sigma^2_{2003} + error^2_{2006,2009} + 3error^2_{2003,2006} - 2\mathrm{cov}\_2003_{2006,2009}
\end{aligned}
$$

9.31

$$
\mathrm{SE}(diff_{2006,2009} - diff_{2003,2006}) = \sqrt{4\sigma^2_{2006} + \sigma^2_{2009} + \sigma^2_{2003} + error^2_{2006,2009} + 3error^2_{2003,2006} - 2\mathrm{cov}\_2003_{2006,2009}}
$$

For example, consider Greece in 2003 to 2009, where we have for Mathematics in 2003 the mean performance 445 with standard error 3.9, 459 and 3.0 in 2006 and 466 and 3.9 in 2009.

So here the difference of the trends is (466-459)-(459-445) =-7. The standard error, using link errors and covariances from Chapter 12, will be

$$
\begin{aligned}
SE &= \sqrt{4(3.0)^2 + (3.9)^2 + (3.9)^2 + (1.33)^2 + 3(1.35)^2 - 2(2.340)} \\
&= \sqrt{68.97} \\
&= 8.31
\end{aligned}
$$

So the test statistic in this case is -7/8.31 = .84, which is not significant at the 5% level. In this case, although the link errors reduce the size of the statistic, the standard errors of the mean are relatively so large that the result of the test would not be changed if they were ignored.

In PISA a common transformation has been estimated, from the link items, and this transformation is applied to all participating countries. It follows that any uncertainty that is introduced through the linking is common to all students and all countries. Thus, for example, suppose the *unknown* linking error (between PISA 2006 and PISA 2009) in reading resulted in an over-estimation of student scores by two points on the PISA 2006 scale. It follows that every student's score will be over-estimated by two score points. This over-estimation will have effects on certain, but not all, summary statistics computed from the PISA 2009 data. For example, consider the following:

- Each country's mean will be over-estimated by an amount equal to the link error, in our example this is two score points.
- The mean performance of any subgroup will be over-estimated by an amount equal to the link error, in our example this is two score points.
- The standard deviation of student scores will not be effected because the over-estimation of each student by a common error does not change the standard deviation.
- The difference between the mean scores of two countries in PISA 2009 will not be influenced because the over-estimation of each student by a common error will have distorted each country's mean by the same amount.
- The difference between the mean scores of two groups (e.g. males and females) in PISA 2009 will not be influenced, because the over-estimation of each student by a common error will have distorted each group's mean by the same amount.
- The difference between the performance of a group of students (e.g. a country) between PISA 2006 and PISA 2009 will be influenced because each student's score in PISA 2006 will be influenced by the error.

- A change in the difference in performance between two groups from PISA 2006 to PISA 2009 will not be influenced. This is because neither of the components of this comparison, which are differences in scores in 2009 and 2006 respectively, is influenced by a common error that is added to all student scores in PISA 2009.

In general terms, the linking error need only be considered when comparisons are being made between results from different PISA data collections, and then usually only when group means are being compared.

Link error for other types of comparisons of student performance

The link error for other comparisons of performance does not have a straightforward theoretical solution as does the link error for comparison between two PISA assessments. The link error between two PISA assessments, described above, can be used, however, to empirically estimate the magnitude of the link error for the comparison of the percentage of students in a particular proficiency level or the magnitude of the link error associated with the estimation of the annualised and curvilinear change.

The empirical estimation of these link errors uses the assumption that the magnitude of the link error follows a normal distribution with mean 0 and a standard deviation equal to the link error or comparisons of performance between PISA 2012 and previous assessments. From this distribution, 500 errors are drawn and added to the first plausible value for each assessment prior to 2012. The estimate of interest (change in the percentage of students in a particular proficiency level or the annualised change) is calculated for each of the 500 replicates. The standard deviation of these 500 estimates is then used as the link error for the annualised change, the quadratic change, and the change in the percentage of students scoring in a particular proficiency level. For further details on these link errors, see OECD (2014), *PISA 2012 Results: What Students Know and Can Do, Student Performance in Mathematics, Reading and Science (Volume I, Revised edition)*, Annex A5.

Note

1. The value M should be large. For PISA we have used 2000.

2. This is a fictitious item.

3.The samples used were simple random samples stratified by the explicit strata used in each country. Students who responded to the UH booklet were not included in this process.

References

Adams, R.J., M. Wilson and **W.C. Wang** (1997), "The Multidimensional Random Coefficients Multinomial Logit Model", *Applied Psychological Measurement*, No. 21, pp. 1-23.

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

Beaton, A.E. (1987), *Implementing the New Design: The NAEP 1983-84 Technical Report* (Rep. No. 15-TR-20), Educational Testing Service, Princeton, NJ.

Ebel, R.L. and **D.A. Frisbie** (1986), *Essentials of Education Measurement*, Prentice Hall, Englewood Cliffs, New Jersey.

Macaskill, G., R.J. Adams and **M.L. Wu** (1998), "Scaling Methodology and Procedures for the Mathematics and Science Literacy, Advanced Mathematics and Physics Scale", in M. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study*, *Technical Report Volume 3: Implementation and Analysis*, Boston College, Chestnut Hill, MA.

Masters, G.N. (1982), "A Rasch Model for Partial Credit Scoring", *Psychometrika*, No. 47(2), pp. 149-174.

Mislevy, R.J. (1991), "Randomization-based Inference about Latent Variables from Complex Samples", *Psychometrika*, No. 56, pp. 177-196.

Mislevy, R.J., A. Beaton, B.A. Kaplan and **K. Sheehan** (1992), "Estimating Population Characteristics from Sparse Matrix Samples of Item Responses", *Journal of Educational Measurement*, No. 29(2), pp. 133-161.

Mislevy, R.J. and **K.M. Sheehan** (1987), “Marginal Estimation Procedures”, in A.E. Beaton (ed.), *The NAEP 1983-84 Technical Report, National Assessment of Educational Progress*, Educational Testing Service, Princeton, pp. 293-360.

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

OECD (2014), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science*, PISA, OECD Publishing, Paris. *http://dx.doi.org/10.1787/9789264208780-en*

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

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

10

Data Management Procedures

INTRODUCTION

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

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

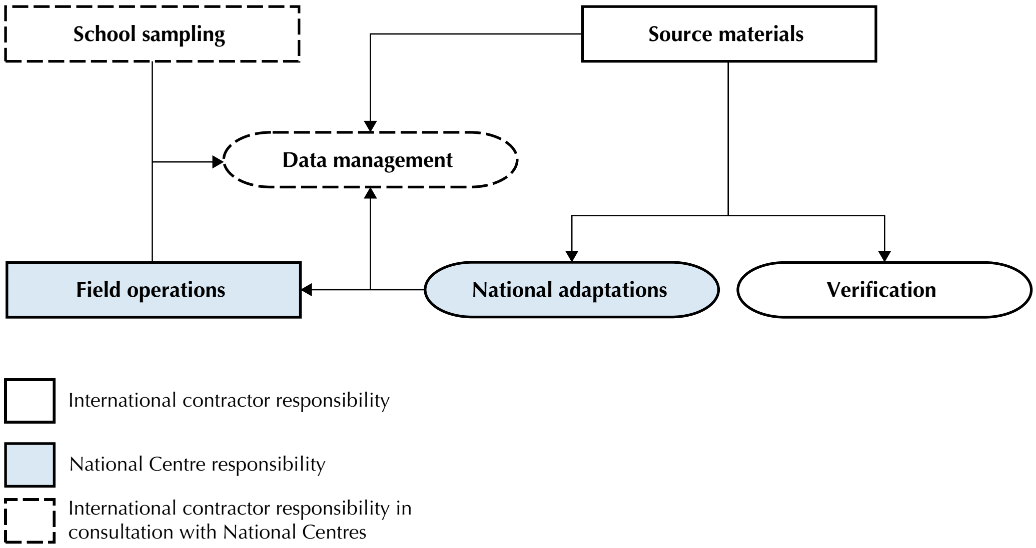
As well as planned variations in the data collected at the national level, the possibility exists for unplanned and unintended variations finding their way into the instruments. Data prepared by national data teams can be corrupted or inaccurate as a result of a number of unintended sources of error. PISA data management procedures are designed to minimise the likelihood of errors occurring, to identify instances where errors may have occurred, and to correct such errors wherever it is possible to do so before the data are finalised. The easiest way to deal with ambiguous or incorrect data would be to delete the whole data record containing values that may be incorrect. However, this should be avoided where possible since the deleted data records results in a decrease in the country's response rate. This chapter will therefore also describe those aspects of data management that are directed at identifying and correcting errors. These procedures applied for both the pencil and paper and computer-delivered components of PISA 2012.

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 contractor, and some are negotiated between the two.

Data management procedures must be shaped to suit the particular cognitive test instruments and background questionnaire instruments used in each participating country. Hence the source materials provided by the international contractor, the national adaptation of those instruments, and the international verification of national versions of all

■ Figure 10.1 ■

Data management in relation to other parts of PISA



instruments must all be reflected in the data management procedures. Data management procedures must also be informed by the outcomes of PISA sampling procedures. The procedures must reliably link data to the students from whom they came. Finally, the test operational procedures that are implemented by each National Centre, and in each test administration session, must be directly related to the data management procedures.

Figure 10.2 illustrates the sequence of major data management tasks in PISA, and shows the division of responsibilities between National Centres, the international contractor, 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, the international contractor provides the data management software *KeyQuest* to all National Centres. *KeyQuest* is a 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 2012.

KeyQuest was preconfigured with all the PISA 2012 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 international contractor approval of proposed national adaptations.

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

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

Once the international contractor 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 the international contractor of any inconsistencies remaining in the data. Further recoding is made after the requests from the National Centres are reviewed. At the same time sampling and tracking data is submitted and analysed and any consequential data recoding is implemented. At that stage the database is regarded as final, and ready for submission to the OECD.

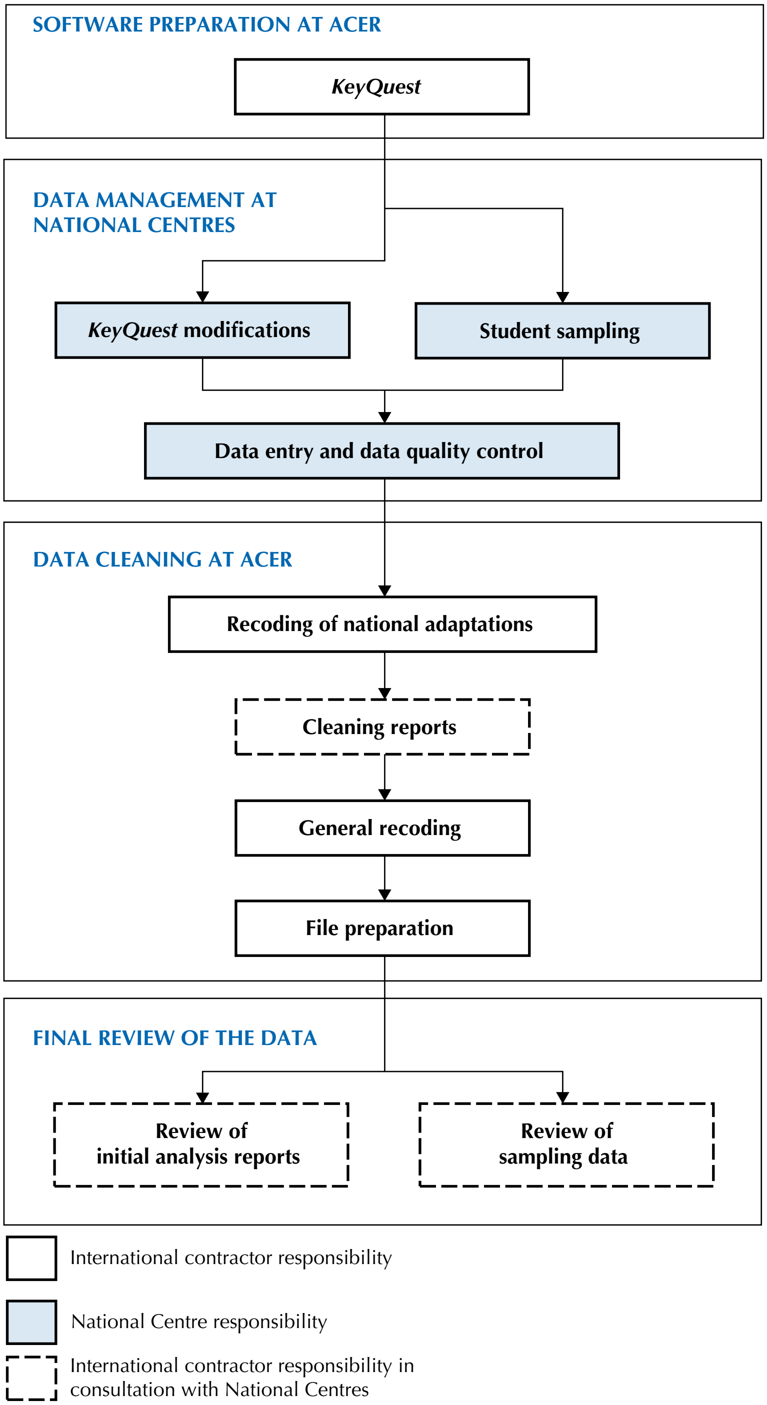
DATA MANAGEMENT AT THE NATIONAL CENTRE

National modifications to the database

PISA's aim is to generate comparable international data from all participating countries, based on a common set of test instruments. However, it is an international study that includes countries with widely differing educational systems and cultural particularities. Due to this diversity, some instrument adaptation is required. Hence verification by the international contractor of national adaptations is crucial (see Chapter 5). After adaptations to the international PISA instruments are agreed upon, the corresponding modifications in *KeyQuest* are made by National Centres.

■ Figure 10.2 ■

Major data management stages in PISA



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* and assigns one of the rotated forms of test booklets and questionnaires to each sampled student. For those countries that participated in the computer-based assessment option of PISA 2012, *KeyQuest* also samples

students from within the sample of students selected for the paper-based assessment and assigns one of the rotated computer-based forms to each sub-sampled student. For countries participating in the International Option of Financial Literacy (FL), the number of students to be sampled for paper-based assessment was increased in each sampled school so as to also achieve the required student sample size for FL. The extra students then assigned FL booklets. Thus, sets of students selected for FL and for main paper-based assessment do not overlap.

In the process of sampling, *KeyQuest* uses the study programme table, which defines the different study programmes operating in sampled schools and enables conversion of national variations into consistent international codes, and the sampling form designed for *KeyQuest,* which summarises all relevant school-level information required to facilitate selection of the school sample. These were agreed with the National Centres through a negotiation process managed through and recorded on the international contractor's website MyPISA (*http://mypisa.acer.edu.au*) and imported into *KeyQuest*.

Critical information about sampled students is captured in the Student Tracking Form in *KeyQuest*. It includes the information used in computing weights, exclusion rates, and participation rates. Other tracking instruments used in *KeyQuest* included the session report form which is used to identify the language of test for each student. The date of the testing session that the student attended is obtained from the session report and used in conjunction with the date of birth of the student entered on the tracking form to calculate the age of the student at the time of testing.

Data entry quality control

The national adaptation and student sampling tasks are performed by staff at each National Centre before testing. After testing, data entry and the running of *KeyQuest* validity reports are carried out by the National Centres.

Validation rules

During data entry, *KeyQuest* captures certain 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 (Missing) was used if none of the choices was circled and code 8 (Invalid) if two or more choices were circled. Finally code 7 (Not Applicable) was reserved for cases when a student was unable to provide a response through no fault of their own, such as when a poorly printed item presented to the student was illegible. Another example is a continuous variable for which it is reasonable to expect any answer from a student in the range from 0 to 200. The particular rule for this question would be to allow entry of values between 0 and 200 (inclusive), which is the range of valid responses for this item. It also allows entry of 9 999 which is used, in this case, to indicate missing data, 9 998 to indicate an invalid response and 9 997 to indicate that the question was not administered. The inbuilt validation rules ensured that no other codes could be entered.

Key violations

Furthermore, *KeyQuest* was programmed to prevent key violations. That is, *KeyQuest* was programmed to prevent the duplication of so called "keys", which are a combination of identifier codes. For example, a data record with the same combination of region, stratum and school identifiers could not be entered twice in the School Questionnaire instrument. A data record with a student ID that does not exist in Student Tracking Form could not be saved at data entry or imported into *KeyQuest* for user-assessable instrument. For more detailed information please refer to PISA 2012 *Data Management Manual.*[1]

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 2012 to keep a count of all records entered by each data entry operator and the time required to enter them. The international contractor recommended for all countries to make use of these procedures for quality assurance purposes during data entry.

Double coding of occupational data

Another optional procedure for PISA 2012 was the double coding of occupational data. The double coding procedure allowed National Centres to check the validity of the data, as well as allowing identification of the areas where

supplementary coding tools could be improved. The main coding tool was the *International Standard Classification of Occupations: ISCO-08* (ILO, 2008)[2] with a small number of additional codes, described in the PISA 2012 *Data Management Manual*. 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. Following this 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 included making improvements to the coding instructions (if the same error was systematically produced by different coders), and/or providing further training for coders that were making more errors than others. Finally, the international contractor expected all discrepancies printed on the report to be resolved before the data were submitted.

The National Centres that participated in this option commented on the usefulness of the procedures for training of the coding staff. The possibilities for the international contractor to conduct a comprehensive analysis of these data were limited due to language constraints. However, one result that was observed was that those countries that required their coders to enter a word description as well as a four-digit code had fewer discrepancies than those that required only a four-digit code. This led to a reinforcement of the ILO (International Labour Organization) recommendation that procedures should involve entering occupation descriptions first and then coding them, rather than coding directly from the questionnaires.

Validity reports

After data entry was completed, National Centres were required to generate validity reports from *KeyQuest* and to resolve all discrepancies listed on these reports before submitting data to the international contractor.

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

- comparison between tracking instruments and sampling verification (tracking instruments, sampling verification);
- data verification within tracking instruments (tracking instruments specific checks);
- comparison of the questionnaire and tracking data (Student Questionnaire - Student Tracking Form specific checks, identity checks questionnaires, identity checks occupation);
- comparison of the identification variables in the test data (identity checks booklets, identity checks CBA - computer-based assessment); and
- verification of the reliability data (reliability checks).

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

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

Due to the complexity and significant number of the validity reports, a validity report checklist was designed. More details about the validity reports can be found in the PISA 2012 *Data Management Manual*.

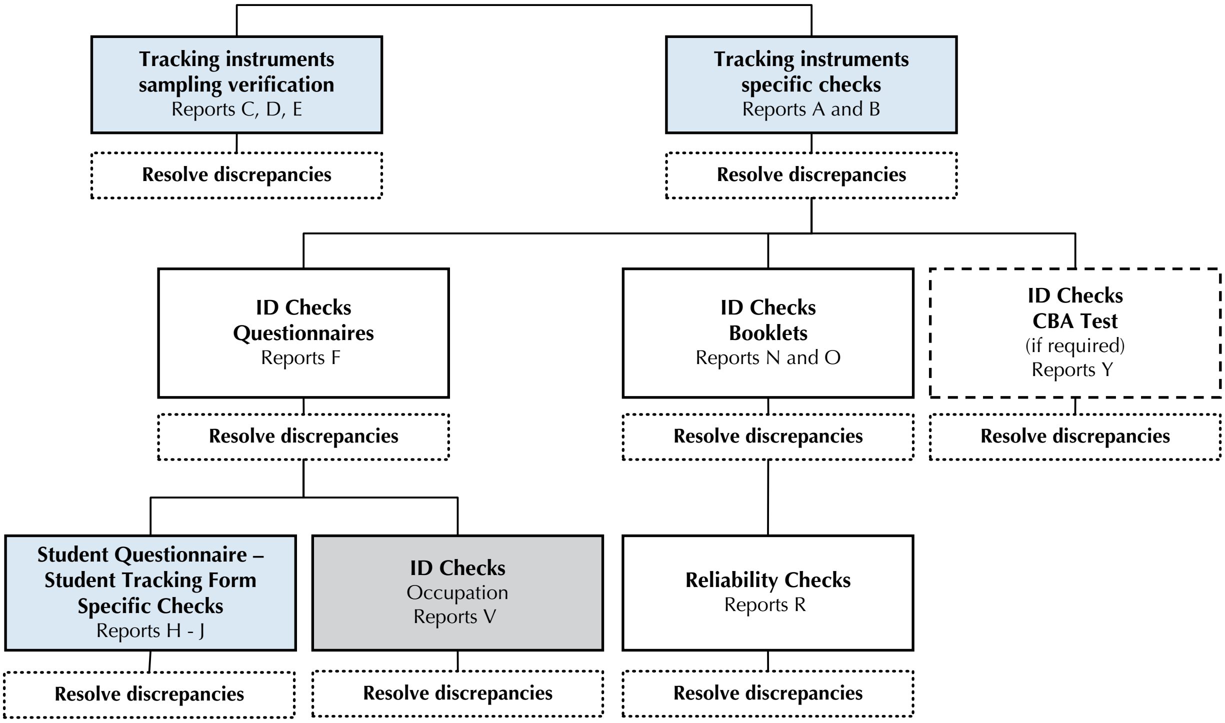
DATA CLEANING BY THE INTERNATIONAL CONTRACTOR

Recoding of national adaptations

When data were submitted by National Centres, 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. After data has been submitted, additional checks were undertaken. During that process of data cleaning, queries, if necessary, are sent to

■ Figure 10.3 ■

Validity Reports: General hierarchy



National Centres, along with reports containing data cleaning results. National Centres clarified all inconsistencies in their database. The national data sets were continuously updated and re-cleaned, to be consistent with information provided by the National Centres. For example, if a variable had been added to a questionnaire, the questionnaire adaptation sheet (which should have been agreed between the National Centre and international contractor before the modification was introduced into *KeyQuest*) was checked to find out whether this national variable required recoding into a 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, necessary recodes for the submitted data to ensure they fit the international structure were performed. 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 international contractor 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. The first step for one or both data cleaners was to check that discrepancies identified in Validity Reports had been resolved. Data cleaners checked all explanations provided in the Validity Reports and in the Item Information for Cleaning document (provided by National Centres at the data submission stage) and, if necessary, requested additional clarification from the relevant National Centres.

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.

Computer-based assessment data

Countries that participated in the computer-based assessment option in PISA 2012 could elect to administer either the problem solving core component only or the problem solving core component plus the computer-based assessment of literacies (mathematics and reading) component. For countries that participated in the computer-based assessment option, the data files constructed from the test delivery and online coding systems were introduced into the data cleaning procedures. Student IDs from the data were checked against student IDs from the paper-based test data (from *KeyQuest*), although data from the computer-based assessment were retained even for those students who had not participated in the PISA paper-based test.

Cleaning reports

During the process of data cleaning, cleaning reports containing the results of the checking procedures were progressively sent to National Centres, with requests to clarify any inconsistencies in their database. The national data sets were then updated according to the information provided by the National Centre.

Many of the cleaning reports were designed to double check the *KeyQuest* validity reports run by the National Centres. If the data had been cleaned correctly at the National Centre, the cleaning reports would either not contain any records or would only list records that had been already explained in the *KeyQuest* validity reports. These cleaning reports were sent only to those countries whose data required additional cleaning.

However, there were additional checks that could not be conducted by the National Centres. For example, inconsistencies within the questionnaires could be checked only after the questionnaire data had been recoded back into the international format. Such cleaning reports were sent to all National Centres.

General re-coding

After receiving all cleaning reports from the National Centres and implementing the agreed corrections recommended in these reports, the international contractor 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 the international contractor about a mistranslation or misprint for an item in the national version of a cognitive booklet, then the data for this item were recoded as “Not Applicable” and were not used in the subsequent analyses.
- If the country deleted a variable in the questionnaire, it was replaced by the “Not Applicable” code.
- If the country changed a variable in the questionnaire in such a way that it could not be recoded into the international format, the international variable was replaced by the “Not Applicable” code.

FINAL REVIEW OF THE DATA

As an outcome of the initial data cleaning, 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 review of national reports provided to each National Centre. After implementation of the corrections made in data cleaning reports and other general recoding, the international contractor sent 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 were summarised in six reports, four of which are described in Chapter 9. For the questionnaires, the reports contained descriptive statistics on every item in the questionnaire.

After review of those initial analysis reports, the National Project Manager (NPM) should have provided information to the international contractor about test items that appeared 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 that was required following initial analyses of the international test and questionnaire data that were prepared for the OECD was introduced into the international data files.

Review of the sampling data

The final data cleaning step of the sampling and tracking data was based on analysis of the national tracking data files. The tracking files and sampling data for each country was checked and analysed and if required requested further recoding were implemented. For example, when a school was regarded as a non-participant because fewer than 25% of students from this school participated in the test, then all students from this school were deleted from the international database. Another example would be a school that was tested outside the permitted test window. All data for students from such a school would also be deleted.

NEXT STEPS IN PREPARING THE INTERNATIONAL DATABASE

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

Notes

1. Available at *http://www.oecd.org/pisa*

2. In this section and throughout the document this edition is called ISCO-08 Manual for short. Available at : *http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm*

Reference

International Labour Organization (ILO) (2007), "ILO plans to support implementation of ISCO-08 in national and regional activities", Paper for discussion by the United Nations Expert Group on International Economic and Social Classifications, New York, April 16-18, 2007.

11

Sampling Outcomes

POPULATION COVERAGE

This chapter reports on PISA sampling outcomes. Details of the sample design are provided in Chapter 4.

Tables 11.1 and 11.2 (by adjudicated regions) show the quality indicators for population coverage and the various pieces of information used to derive them. The following notes explain the meaning of each coverage index and how the data in each column of the table were used.

Coverage Indices 1, 2 and 3 are intended to measure PISA population coverage. Coverage Indices 4 and 5 are intended to be diagnostic in cases where indices 1, 2 or 3 have expected values. Many references are made in this chapter to the various sampling tasks on which National Project Managers (NPMs) documented statistics and other information needed in undertaking the sampling of schools and students. Note that although no comparison is made between the total population of 15-year-olds and the enrolled population of 15-year-old students, generally the enrolled population was expected to be less than or equal to the total population. Occasionally this was not the case due to differing data sources for these two values.

Coverage Index 1: Coverage of the national population, calculated by $P/(P+E) \times$ (ST7b_3/ST7b_1):

- Coverage Index 1 shows the extent to which the weighted participants covered the final target population after all school exclusions. The following bullet points give details of its computation.
- In the preceding expression $P/(P+E)$ broadly represents the coverage proportion due to within school exclusion, and (ST7b_3/ST7b_1) the coverage proportion due to school level exclusion.
- The national population (NP) value, defined by Sampling Task 7b response box [1] and denoted here as ST7b_1 (and in Table 11.1 as the target population) is the population that includes all enrolled 15-year-old students in grades 7 and above in each participating country (with the possibility of small levels of exclusions), based on national statistics. However, the final NP value reflected for each country's school sampling frame might have had some school-level exclusions. The value that represents the population of enrolled 15-year-old students minus those in excluded schools is represented initially by response box [3] on Sampling Task 7b. It is denoted here as ST7b_3. As in PISA 2009, the procedure for PISA 2012 was that very small schools having only one or two PISA-eligible students could not be excluded from the school frame but could be excluded in the field if the school still had only one or two PISA-eligible students at the time of data collection. Therefore, what is noted in Coverage Index 1 as ST7b_3 (and in Table 11.1 as target minus school level exclusions) was a number after accounting for all school level exclusions, which means a number that excludes schools excluded from the sampling frame in addition to those schools excluded in the field. Thus, the term (ST7b_3/ST7b_1) provides the proportion of the NP covered in each country based on national statistics.
- The value $(P+E)$ provides the weighted estimate from the student sample of all PISA-eligible 15-year-olds in each participating country, where P is the weighted estimate of PISA-eligible non-excluded 15-year-old students and E is the weighted estimate of PISA-eligible 15-year-old students that were excluded within schools. Therefore, the term $P/(P+E)$ provides an estimate, based on the student sample, of the proportion of the PISA-eligible 15-year-old population represented by the non-excluded PISA-eligible 15-year-old students.
- The result of multiplying these two proportions together ($P/(P+E)$ and (ST7b_3/ST7b_1)) indicates the overall proportion of the NP covered by the non-excluded portion of the student sample.

Coverage Index 2: Coverage of the national enrolled population, calculated by $P/(P+E) \times$ (ST7b_3/ST7a_2.1):

- Coverage Index 2 shows the extent to which the weighted participants covered the target population of all enrolled students in grades 7 and above.
- The national enrolled population (NEP), defined by Sampling Task 7a response box [2.1] and denoted here as ST7a_2.1 (and as enrolled 15-year-old students in Table 11.1), is the population that includes all enrolled 15-year-old students in grades 7 and above in each participating country, based on national statistics. The final NP, denoted here as ST7b_3 as described above for Coverage Index 1, reflects the 15- year-old population after school-level and other small exclusions. This value represents the population of enrolled 15-year-old students less those in excluded schools.
- The value $(P+E)$ provides the weighted estimate from the student sample of all eligible 15-year-olds in each country, where P is the weighted estimate of PISA-eligible non-excluded 15-year-old students and E is the weighted estimate of PISA-eligible 15-year-old students that were excluded within schools. Therefore, the term $P/(P+E)$ provides an estimate based on the student sample of the proportion of the PISA-eligible 15-year-old population that is represented by the non-excluded PISA-eligible 15-year-old students.

- Multiplying these two proportions together ($P/(P+E)$ and (ST7b_3/ST7a_2.1)) gives the overall proportion of the NEP that was covered by the non-excluded portion of the student sample.

So Coverage Index 1 and Coverage Index 2 will differ when countries have excluded geographical areas or language groups apart from other school level exclusions. In these cases Coverage Index 2 will be less than Coverage Index 1.

Coverage Index 3: Coverage of the national 15-year-old population, calculated by *P*/ST7a_1:

- The national population of 15-year-olds, defined by Sampling Task 7a response box [1] and denoted here as ST7a_1 (and called all 15-year-olds in Table 11.1), is the entire population of 15-year-olds in each country (enrolled and not enrolled), based on national statistics. The value P is the weighted estimate of PISA-eligible non-excluded 15-year-old students from the student sample. Thus (*P*/ST7a_1) indicates the proportion of the national population of 15-year-olds covered by the non-excluded portion of the student sample. It therefore reflects the proportion of 15 year olds excluded or not at school.

Coverage Index 4: Coverage of the estimated school population, calculated by $(P+E)/S$:

- The value $(P+E)$ provides the weighted estimate from the student sample of all PISA-eligible 15-year-old students in each country, where P is the weighted estimate of PISA-eligible non-excluded 15-year-old students and E is the weighted estimate of PISA-eligible 15-year-old students who were excluded within schools.
- The value S is an estimate of the 15-year-old school population in each participating country (called estimate of enrolled students from frame in Table 11.1). This is based on the actual or (more often) approximate number of 15-year-old students enrolled in each school in the sample, prior to contacting the school to conduct the assessment,. The S value is calculated as the sum over all sampled schools of the product of each school's sampling weight and its number of 15-year-old students (*ENR*) as recorded on the school sampling frame.
- Thus, $(P+E)/S$ is the proportion of the estimated school 15-year-old population that is represented by the weighted estimate from the student sample of all PISA-eligible 15-year-old students. It will be influenced by the accuracy of the school sample frame, fluctuations in the target population size and the accuracy of the within school sampling process. Its purpose is to check whether the student sampling has been carried out correctly, and to assess whether the value of S is a reliable measure of the number of enrolled 15-year-olds. This is important for interpreting Coverage Index 5.

Coverage Index 5: Coverage of the school sampling frame population, calculated by *S*/ST7b_3:

- The value (*S*/ST7b_3) is the ratio of the enrolled 15-year-old population, as estimated from data on the school sampling frame, to the size of the enrolled student population, as reported on Sampling Task 7b and adjusted by removing any additional excluded schools in the field. In some cases, this provided a check as to whether the data on the sampling frame gave a reliable estimate of the number of 15-year-old students in each school. In other cases, however, it was evident that ST7b_3 had been derived using data from the sampling frame by the NPM, so that this ratio may have been close to 1.0 even if enrolment data on the school sampling frame were poor. Under such circumstances, Coverage Index 4 would differ noticeably from 1.0, and the figure for ST7b_3 would also be inaccurate.

SCHOOL AND STUDENT RESPONSE RATES

Tables 11.3 to 11.8 present school and student-level response rates at the national and regional levels.

- Tables 11.3 and 11.4 (by adjudicated regions) indicates the rates calculated by using only original schools and no replacement schools.
- Tables 11.5 and 11.6 (by adjudicated regions) indicates the improved response rates when first and second replacement schools were accounted for in the rates.
- Tables 11.7 and 11.8 (by adjudicated regions) indicates the student response rates among the full set of participating schools.

[Part 1/2]

Table 11.1 PISA target populations and samples

		All 15-year-olds	Enrolled 15-year-olds	Target population	School-level exclusions	Target minus school level exclusions	School level exclusion rate (%)	Estimation of enrolled students from frame	Number participating students	Weighted number of participating students	Number of excluded students
OECD	Australia	291967	288159	288159	5 702	282 457	1.98	274 432	17 774	250 779	505
	Austria	93537	89073	89073	106	88 967	0.12	88 967	4 756	82 242	46
	Belgium	123469	121493	121209	1 324	119 885	1.09	119 019	9 690	117 912	39
	Canada	417873	409453	404767	2 936	401 831	0.73	396 757	21 548	348 070	1796
	Chile	274803	252733	252625	2 687	249 938	1.06	239 429	6 857	229 199	18
	Czech Republic	96946	93214	93214	1 577	91 637	1.69	88 884	6 535	82 101	15
	Denmark	72310	70854	70854	1 965	68 889	2.77	71 015	7 481	65 642	368
	Estonia	12649	12438	12438	442	11 996	3.55	12 046	5 867	11 634	143
	Finland	62523	62195	62195	523	61 672	0.84	60 323	8 829	60 047	225
	France	792983	755447	755447	27 403	728 044	3.63	728 401	5 682	701 399	52
	Germany	798136	798136	798136	10 914	787 222	1.37	753 179	5 001	756 907	8
	Greece	110521	105096	105096	1 364	103 732	1.30	102 087	5 125	96 640	136
	Hungary	111761	108816	108816	1 725	107 091	1.59	101 751	4 810	91 179	27
	Iceland	4505	4491	4491	10	4 481	0.22	4 424	3 508	4 169	155
	Ireland	59296	57979	57952	0	57 952	0.00	57 711	5 016	54 010	271
	Israel	118953	113278	113278	2 784	110 494	2.46	109 326	6 061	107 745	114
	Italy	605490	566973	566973	8 498	558 475	1.50	536 921	38 142	521 288	741
	Japan	1241786	1214756	1214756	26 099	1 188 657	2.15	1 175 794	6 351	1 128 179	0
	Korea	687104	672101	672101	3 053	669 048	0.45	662 510	5 033	603 632	17
	Luxembourg	6187	6082	6082	151	5 931	2.48	5 931	5 260	5 523	357
	Mexico	2114745	1472875	1472875	7 307	1 465 568	0.50	1 442 242	33 806	1 326 025	58
	Netherlands	194000	193190	193190	7 546	185 644	3.91	185 468	4 460	196 262	27
	New Zealand	60940	59118	59118	579	58 539	0.98	58 676	5 248	53 414	255
	Norway	64917	64777	64777	750	64 027	1.16	63 653	4 686	59 432	278
	Poland	425597	410700	410700	6 900	403 800	1.68	402 116	5 662	379 275	212
	Portugal	108728	127537	127537	0	127 537	0.00	128 129	5 722	96 034	124
	Slovak Republic	59723	59367	59367	1 480	57 887	2.49	57 353	5 737	54 486	29
	Slovenia	19471	18935	18935	115	18 820	0.61	18 680	7 229	18 303	84
	Spain	423444	404374	404374	2 031	402 343	0.50	403 999	25 335	374 266	959
	Sweden	102087	102027	102027	1 705	100 322	1.67	99 726	4 739	94 988	201
	Switzerland	87200	85239	85239	2 479	82 760	2.91	83 450	11 234	79 679	256
	Turkey	1266638	965736	965736	10 387	955 349	1.08	945 357	4 848	866 681	21
	United Kingdom	738 066	745 581	745 581	19 820	725 761	2.66	705 011	12 659	688 236	486
	United States	3985714	4074457	4074457	41 142	4 033 315	1.01	3 945 575	6 111	3 536 153	319
Partners	Albania	76910	50157	50157	56	50 101	0.11	49 632	4 743	42 466	1
	Argentina	684879	637603	637603	3 995	633 608	0.63	606 069	5 908	545 942	12
	Brazil	3574928	2786064	2786064	34 932	2 751 132	1.25	2 745 045	20 091	2 470 804	44
	Bulgaria	70188	59684	59684	1 437	58 247	2.41	57 574	5 282	54 255	6
	Colombia	889729	620422	620422	4	620 418	0.00	612 605	11 173	560 805	23
	Costa Rica	81489	64326	64326	0	64 326	0.00	64 920	4 602	40 384	2
	Croatia	48155	46550	46550	417	46 133	0.90	45 636	6 153	45 502	91
	Cyprus[1, 2]	9956	9956	9955	128	9 827	1.29	9 821	5 078	9 650	157
	Hong Kong-China	84200	77864	77864	813	77 051	1.04	76 589	4 670	70 636	38
	Indonesia	4174217	3599844	3544028	8 039	3 535 989	0.23	2 950 696	5 622	2 645 155	2
	Jordan	129492	125333	125333	141	125 192	0.11	119 147	7 038	111 098	19
	Kazakhstan	258716	247048	247048	7 374	239 674	2.98	239 767	5 808	208 411	25
	Latvia	18789	18389	18375	655	17 720	3.56	17 488	5 276	16 054	14
	Liechtenstein	417	383	383	1	382	0.26	382	293	314	13
	Lithuania	38524	35567	35567	526	35 041	1.48	34 614	4 618	33 042	130
	Macao-China	6600	5416	5416	6	5 410	0.11	5 410	5 335	5 366	3
	Malaysia	544302	457999	457999	225	457 774	0.05	455 543	5 197	432 080	7
	Montenegro	8600	8600	8600	18	8 582	0.21	8 540	4 744	7 714	4
	Peru	584294	508969	508969	263	508 706	0.05	514 574	6 035	419 945	8
	Qatar	11667	11532	11532	202	11 330	1.75	11 340	10 966	11 003	85
	Romania	146243	146243	146243	5 091	141 152	3.48	139 597	5 074	140 915	0
	Russian Federation	1272632	1268814	1268814	17 800	1 251 014	1.40	1 243 564	6 418	1 172 539	69
	Serbia	80089	75870	74272	1 987	72 285	2.67	72 819	4 684	67 934	10
	Shanghai-China	108 056	90796	90796	1 252	89 544	1.38	89 832	6 374	85 127	8
	Singapore	53637	52163	52163	293	51 870	0.56	51 687	5 546	51 088	33
	Chinese Taipei	328356	328336	328336	1 747	326 589	0.53	324 667	6 046	292 542	44
	Thailand	982080	784897	784897	9 123	775 774	1.16	772 654	6 606	703 012	12
	Tunisia	132313	132313	132313	169	132 144	0.13	130 141	4 407	120 784	5
	United Arab Emirates	48824	48446	48446	971	47 475	2.00	46 748	11 500	40 612	11
	Uruguay	54638	46442	46442	14	46 428	0.03	46 009	5 315	39 771	15
	Viet Nam	1717996	1091462	1091462	7 729	1 083 733	0.71	1 068 462	4 959	956 517	1

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.1 PISA target populations and samples

		Weighted number of excluded students	Number of ineligible students	Weighted number of ineligible students	Within school exclusion rate (%)	Overall exclusion rate (%)	Percentage ineligible / withdrawn	Coverage Index 1	Coverage Index 2	Coverage Index 3	Coverage Index 4	Coverage Index 5
OECD	Australia	5 282	1185	9 815	2.06	4.00	3.83	0.96	0.96	0.86	0.93	0.97
	Austria	1 011	131	2 042	1.21	1.33	2.45	0.99	0.99	0.88	0.94	1.00
	Belgium	367	179	2 183	0.31	1.40	1.85	0.99	0.98	0.95	0.99	0.99
	Canada	21 013	1287	16 938	5.69	6.38	4.59	0.94	0.93	0.83	0.93	0.99
	Chile	548	150	5 417	0.24	1.30	2.36	0.99	0.99	0.83	0.96	0.96
	Czech Republic	118	55	643	0.14	1.83	0.78	0.98	0.98	0.85	0.93	0.97
	Denmark	2 381	104	671	3.50	6.18	0.99	0.94	0.94	0.91	0.96	1.03
	Estonia	277	29	53	2.33	5.80	0.45	0.94	0.94	0.92	0.99	1.00
	Finland	653	66	322	1.08	1.91	0.53	0.98	0.98	0.96	1.01	0.98
	France	5 828	184	20 613	0.82	4.42	2.91	0.96	0.96	0.88	0.97	1.00
	Germany	1 302	56	7 805	0.17	1.54	1.03	0.98	0.98	0.95	1.01	0.96
	Greece	2 304	64	1 074	2.33	3.60	1.09	0.96	0.96	0.87	0.97	0.98
	Hungary	928	43	760	1.01	2.58	0.82	0.97	0.97	0.82	0.91	0.95
	Iceland	156	27	27	3.60	3.81	0.63	0.96	0.96	0.93	0.98	0.99
	Ireland	2 524	70	681	4.47	4.47	1.20	0.96	0.95	0.91	0.98	1.00
	Israel	1 884	36	653	1.72	4.13	0.60	0.96	0.96	0.91	1.00	0.99
	Italy	9 855	842	12 387	1.86	3.33	2.33	0.97	0.97	0.86	0.99	0.96
	Japan	0	6	885	0.00	2.15	0.08	0.98	0.98	0.91	0.96	0.99
	Korea	2 238	83	9 989	0.37	0.82	1.65	0.99	0.99	0.88	0.91	0.99
	Luxembourg	357	14	14	6.07	8.40	0.24	0.87	0.87	0.85	0.99	1.00
	Mexico	3 247	2877	107 100	0.24	0.74	8.06	0.99	0.99	0.63	0.92	0.98
	Netherlands	1 056	31	1 115	0.54	4.42	0.56	0.96	0.96	1.01	1.06	1.00
	New Zealand	2 030	186	1 626	3.66	4.61	2.93	0.95	0.95	0.88	0.94	1.00
	Norway	3 133	31	353	5.01	6.11	0.56	0.94	0.94	0.92	0.98	0.99
	Poland	11 566	28	1 593	2.96	4.59	0.41	0.95	0.95	0.89	0.97	1.00
	Portugal	1 560	237	3 151	1.60	1.60	3.23	0.98	0.98	0.88	0.76	1.00
	Slovak Republic	246	29	266	0.45	2.93	0.49	0.97	0.97	0.91	0.95	0.99
	Slovenia	181	94	195	0.98	1.58	1.05	0.98	0.98	0.94	0.99	0.99
	Spain	14 931	369	7 537	3.84	4.32	1.94	0.96	0.96	0.88	1.01	0.96
	Sweden	3 789	44	869	3.84	5.44	0.88	0.95	0.95	0.93	0.99	0.99
	Switzerland	1 093	148	1 028	1.35	4.22	1.27	0.96	0.96	0.91	0.97	1.01
	Turkey	3 684	287	50 864	0.42	1.49	5.84	0.99	0.99	0.68	0.92	0.99
	United Kingdom	20 173	517	19 599	2.85	5.43	2.77	0.95	0.95	0.93	1.00	0.97
	United States	162 194	269	133 855	4.39	5.35	3.62	0.95	0.95	0.89	0.94	0.98
Partners	Albania	10	117	909	0.02	0.14	2.14	1.00	1.00	0.55	0.86	0.99
	Argentina	641	203	16 350	0.12	0.74	2.99	0.99	0.99	0.80	0.90	0.96
	Brazil	4 900	1625	149 383	0.20	1.45	6.03	0.99	0.99	0.69	0.90	1.00
	Bulgaria	80	49	521	0.15	2.55	0.96	0.97	0.97	0.77	0.94	0.99
	Colombia	789	460	24 211	0.14	0.14	4.31	1.00	1.00	0.63	0.92	0.99
	Costa Rica	12	201	1 620	0.03	0.03	4.01	1.00	1.00	0.50	0.62	1.01
	Croatia	627	87	603	1.36	2.24	1.31	0.98	0.98	0.94	1.01	0.99
	Cyprus[1, 2]	200	55	85	2.03	3.29	0.87	0.97	0.97	0.97	1.00	1.00
	Hong Kong-China	518	36	501	0.73	1.76	0.70	0.98	0.98	0.84	0.93	0.99
	Indonesia	860	50	22 058	0.03	0.26	0.83	1.00	0.98	0.63	0.90	0.83
	Jordan	304	303	4 485	0.27	0.39	4.03	1.00	1.00	0.86	0.93	0.95
	Kazakhstan	951	49	1 603	0.45	3.43	0.77	0.97	0.97	0.81	0.87	1.00
	Latvia	76	45	128	0.47	4.02	0.79	0.96	0.96	0.85	0.92	0.99
	Liechtenstein	13	6	6	3.97	4.22	1.83	0.96	0.96	0.75	0.86	1.00
	Lithuania	867	40	258	2.56	4.00	0.76	0.96	0.96	0.86	0.98	0.99
	Macao-China	3	29	29	0.06	0.17	0.54	1.00	1.00	0.81	0.99	1.00
	Malaysia	554	197	15 921	0.13	0.18	3.68	1.00	1.00	0.79	0.95	1.00
	Montenegro	8	50	71	0.10	0.31	0.92	1.00	1.00	0.90	0.90	1.00
	Peru	549	337	23 851	0.13	0.18	5.67	1.00	1.00	0.72	0.82	1.01
	Qatar	85	383	383	0.77	2.51	3.46	0.97	0.97	0.94	0.98	1.00
	Romania	0	51	1 403	0.00	3.48	1.00	0.97	0.97	0.96	1.01	0.99
	Russian Federation	11 940	70	12 957	1.01	2.40	1.09	0.98	0.98	0.92	0.95	0.99
	Serbia	136	81	1 050	0.20	2.87	1.54	0.97	0.95	0.85	0.93	1.01
	Shanghai-China	107	71	916	0.13	1.50	1.08	0.98	0.98	0.79	0.95	1.00
	Singapore	315	63	513	0.61	1.17	1.00	0.99	0.99	0.95	0.99	1.00
	Chinese Taipei	2 029	136	5 532	0.69	1.22	1.88	0.99	0.99	0.89	0.91	0.99
	Thailand	1 144	225	27 077	0.16	1.32	3.85	0.99	0.99	0.72	0.91	1.00
	Tunisia	130	271	6 818	0.11	0.24	5.64	1.00	1.00	0.91	0.93	0.98
	United Arab Emirates	37	299	1 231	0.09	2.09	3.03	0.98	0.98	0.83	0.87	0.98
	Uruguay	99	409	2 910	0.25	0.28	7.30	1.00	1.00	0.73	0.87	0.99
	Viet Nam	198	129	23 725	0.02	0.73	2.48	0.99	0.99	0.56	0.90	0.99

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]
Table 11.2 PISA target populations and samples, by adjudicated regions

		All 15-year-olds	Enrolled 15-year-olds	Target population	School-level exclusions	Target minus school level exclusions	School level exclusion rate (%)	Estimation of enrolled students from frame	Number participating students	Weighted number of participating students	Number of excluded students
OECD	Belgium (Flemish community)	71 588	68 823	68 823	1 764	67 059	2.32	67 031	5 970	66 734	22
	Spain (Andalusia)	86 402	82 526	82 526	256	82 270	0.31	82 759	1 439	75 819	52
	Spain (Aragon)	11 276	11 225	11 225	60	11 165	0.53	11 024	1 393	9 988	55
	Spain (Asturias)	7 180	7 000	7 000	44	6 956	0.63	7 014	1 611	7 125	43
	Spain (Balearic Islands)	10 369	9 198	9 198	40	9 158	0.43	9 156	1 435	8 385	122
	Spain (Basque Country)	16 955	16 943	16 943	9	16 934	0.05	16 854	4 740	16 146	123
	Spain (Cantabria)	4 507	4 448	4 448	11	4 437	0.25	4 474	1 523	4 334	22
	Spain (Castile and Leon)	20 155	19 962	19 962	120	19 842	0.38	19 737	1 592	18 422	62
	Spain (Catalonia)	66 016	64 079	64 079	629	63 450	0.98	63 774	1 440	55 982	90
	Spain (Extremadura)	11 160	11 048	11 048	47	11 001	0.43	10 978	1 537	10 407	49
	Spain (Galicia)	20 994	19 175	19 175	26	19 149	0.14	19 095	1 544	18 316	53
	Spain (La Rioja)	2 733	2 606	2 606	10	2 596	0.38	2 596	1 533	2 568	72
	Spain (Madrid)	57 304	53 159	53 159	366	52 793	0.69	53 017	1 543	48 880	38
	Spain (Murcia)	15 247	14 502	14 502	29	14 473	0.20	14 207	1 379	13 185	67
	Spain (Navarre)	5 584	5 517	5 517	18	5 499	0.33	5 478	1 530	5 245	48
	United Kingdom (Scotland)	61863	60769	60769	1063	59706	1.75	59 727	2 945	53 898	115
	United States (Connecticut)[1]	48 970	49 996	49 996	7 983	210 352	3.66	43 310	1 697	37 053	73
	United States (Florida)[1]	226 434	218 335	218 335	290	49 706	0.58	191 020	1 896	174 318	98
	United States (Massachusetts)[1]	81 580	84 320	84 320	849	83 471	1.01	69 336	1 723	56 991	71
Partners	Argentina (CABA)	36 183	36 694	36 694	131	36 563	0.36	35 784	1 336	33 009	6
	Russian Federation (Perm Territory)	28 437	28 352	28 352	740	27 612	2.61	28 260	1 761	27 776	6
	United Arab Emirates (Abu Dhabi)	16 974	16 871	16 871	186	16 685	1.10	16 385	3 163	15 715	2
	United Arab Emirates (Dubai)	14 515	14 437	14 437	621	13 816	4.30	13 654	4 974	10 758	8

1. Only public.

[Part 2/2]
Table 11.2 PISA target populations and samples, by adjudicated regions

		Weighted number of excluded students	Number of ineligible students	Weighted number of ineligible students	Within school exclusion rate (%)	Overall exclusion rate (%)	Percentage ineligible / withdrawn	Coverage Index 1	Coverage Index 2	Coverage Index 3	Coverage Index 4	Coverage Index 5
OECD	Belgium (Flemish community)	245	84	870	0.37	2.92	1.30	0.97	0.97	0.93	1.00	1.00
	Spain (Andalusia)	2 421	80	3 735	3.09	3.40	4.77	0.97	0.97	0.88	0.95	1.01
	Spain (Aragon)	367	42	310	3.54	4.06	2.99	0.96	0.96	0.89	0.94	0.99
	Spain (Asturias)	164	11	57	2.25	2.86	0.78	0.97	0.97	0.99	1.04	1.01
	Spain (Balearic Islands)	632	24	121	7.01	7.41	1.34	0.93	0.93	0.81	0.98	1.00
	Spain (Basque Country)	425	13	41	2.57	2.62	0.25	0.97	0.97	0.95	0.98	1.00
	Spain (Cantabria)	58	28	69	1.31	1.56	1.58	0.98	0.98	0.96	0.98	1.01
	Spain (Castile and Leon)	693	4	44	3.63	4.21	0.23	0.96	0.96	0.91	0.97	0.99
	Spain (Catalonia)	3 250	24	853	5.49	6.42	1.44	0.94	0.94	0.85	0.93	1.01
	Spain (Extremadura)	293	20	121	2.74	3.15	1.13	0.97	0.97	0.93	0.97	1.00
	Spain (Galicia)	579	11	153	3.06	3.20	0.81	0.97	0.97	0.87	0.99	1.00
	Spain (La Rioja)	103	15	27	3.86	4.23	1.03	0.96	0.96	0.94	1.03	1.00
	Spain (Madrid)	1 119	13	372	2.24	2.91	0.74	0.97	0.97	0.85	0.94	1.00
	Spain (Murcia)	530	55	430	3.87	4.06	3.13	0.96	0.96	0.86	0.97	0.98
	Spain (Navarre)	153	11	32	2.83	3.15	0.59	0.97	0.97	0.94	0.99	1.00
	United Kingdom (Scotland)	1 716	210	3 208	3.09	4.78	5.77	0.95	0.95	0.87	0.93	1.00
	United States (Connecticut)[1]	1 367	87	1 582	3.56	4.12	4.12	0.96	0.96	0.76	0.89	0.87
	United States (Florida)[1]	8 802	66	5 582	4.81	8.29	3.05	0.92	0.92	0.77	0.96	0.91
	United States (Massachusetts)[1]	2 022	77	2 178	3.43	4.40	3.69	0.96	0.96	0.70	0.85	0.83
Partners	Argentina (CABA)	76	19	401	0.23	0.59	1.21	0.99	0.99	0.91	0.92	0.98
	Russian Federation (Perm Territory)	88	7	119	0.32	2.92	0.43	0.97	0.97	0.98	0.99	1.02
	United Arab Emirates (Abu Dhabi)	8	85	446	0.05	1.15	2.84	0.99	0.99	0.93	0.96	0.98
	United Arab Emirates (Dubai)	11	137	422	0.10	4.40	3.92	0.96	0.96	0.74	0.79	0.99

1. Only public.

For calculating school response rates before replacement, the numerator consisted of all original sample schools with enrolled age-eligible students who participated (i.e. assessed a sample of PISA-eligible students, and obtained a student response rate of at least 50%). The denominator consisted of all the schools in the numerator, plus those original sample schools with enrolled age-eligible students that either did not participate or failed to assess at least 50% of PISA-eligible sample students. Schools that were included in the sampling frame, but were found to have no age-eligible students, or which were excluded in the field were omitted from the calculation of response rates. Replacement schools do not figure in these calculations.

Table 11.3 Response rates before school replacement

		Weighted school participation rate before replacement (%) (SCHRRW1)	Weighted number of responding schools (weighted also by enrollment) (NUMW1)	Weighted number of schools sampled (responding + non-responding) (weighted also by enrollment) (DENW1)	Unweighted school participation rate before replacement (%) (SCHRRU1)	Number of responding schools (unweighted) (NUMU1)	Number of responding and non-responding schools (unweighted) (DENU1)
OECD	Australia	97.89	268630.57	274432.06	95.82	757	790
	Austria	100.00	88967.28	88967.28	100.00	191	191
	Belgium	84.43	100481.79	119018.79	83.67	246	294
	Canada	91.28	362178.13	396757.36	91.29	828	907
	Chile	91.89	220008.93	239429.03	89.29	200	224
	Czech Republic	98.15	87238.40	88884.40	98.32	292	297
	Denmark	86.95	61749.44	71014.70	84.97	311	366
	Estonia	100.00	12046.26	12046.26	100.00	206	206
	Finland	99.03	59739.55	60322.79	99.04	310	313
	France	96.58	703458.22	728400.97	96.54	223	231
	Germany	97.71	735943.62	753179.40	97.42	227	233
	Greece	93.16	95107.00	102086.64	91.67	176	192
	Hungary	97.61	99316.84	101751.21	95.19	198	208
	Iceland	99.34	4395.00	4424.00	95.00	133	140
	Ireland	98.70	56962.29	57711.25	98.38	182	185
	Israel	91.05	99542.68	109326.24	89.25	166	186
	Italy	89.09	478317.05	536921.00	89.61	1104	1232
	Japan	86.34	1015198.08	1175793.91	86.50	173	200
	Korea	99.86	661575.41	662509.92	99.36	156	157
	Luxembourg	100.00	5931.00	5931.00	100.00	42	42
	Mexico	91.79	1323815.85	1442242.13	91.61	1431	1562
	Netherlands	75.33	139709.29	185468.01	74.37	148	199
	New Zealand	80.85	47441.43	58676.31	79.19	156	197
	Norway	85.15	54200.68	63652.93	85.10	177	208
	Poland	85.38	343344.29	402116.47	84.57	159	188
	Portugal	95.40	122238.33	128129.44	95.38	186	195
	Slovak Republic	87.50	50181.67	57353.16	85.59	202	236
	Slovenia	98.12	18328.68	18680.14	94.90	335	353
	Spain	99.65	402604.32	403998.99	99.78	902	904
	Sweden	98.92	98645.35	99726.24	98.10	207	211
	Switzerland	94.46	78825.04	83449.65	94.08	397	422
	Turkey	97.49	921643.25	945356.78	97.06	165	170
	United Kingdom	80.06	564438.06	705011.06	86.73	477	550
	United States	67.09	2647252.95	3945574.93	67.15	139	207
Partners	Albania	100.00	49631.96	49631.96	100.00	204	204
	Argentina	95.49	578723.08	606069.17	95.20	218	229
	Brazil	92.74	2545862.67	2745044.76	90.63	803	886
	Bulgaria	99.18	57100.72	57574.24	98.94	186	188
	Colombia	86.61	530553.43	612605.39	88.98	323	363
	Costa Rica	98.94	64234.83	64919.87	98.96	191	193
	Croatia	98.69	45036.68	45636.39	98.17	161	164
	Cyprus[1, 2]	96.58	9485.00	9821.00	89.31	117	131
	Hong Kong-China	78.70	60277.30	76588.65	78.85	123	156
	Indonesia	94.89	2799942.53	2950695.75	94.76	199	210
	Jordan	100.00	119147.04	119147.04	100.00	233	233
	Kazakhstan	100.00	239767.02	239767.02	100.00	218	218
	Latvia	87.89	15370.87	17488.24	87.32	186	213
	Liechtenstein	100.00	382.00	382.00	100.00	12	12
	Lithuania	98.20	33989.27	34613.52	97.69	211	216
	Macao-China	100.00	5410.00	5410.00	100.00	45	45
	Malaysia	100.00	455543.42	455543.42	100.00	164	164
	Montenegro	100.00	8540.46	8540.46	100.00	51	51
	Peru	97.93	503914.74	514574.26	97.94	238	243
	Qatar	99.94	11333.00	11340.00	95.73	157	164
	Romania	100.00	139597.34	139597.34	100.00	178	178
	Russian Federation	100.00	1243563.91	1243563.91	100.00	227	227
	Serbia	90.00	65536.95	72819.24	89.38	143	160
	Shanghai-China	100.00	89831.82	89831.82	100.00	155	155
	Singapore	97.54	50414.86	51687.10	96.59	170	176
	Chinese Taipei	100.00	324667.48	324667.48	100.00	163	163
	Thailand	98.04	757516.07	772654.27	97.92	235	240
	Tunisia	99.30	129228.71	130141.41	99.35	152	153
	United Arab Emirates	99.40	46468.67	46747.80	98.48	453	460
	Uruguay	99.41	45735.66	46009.37	99.44	179	180
	Viet Nam	100.00	1068461.51	1068461.51	100.00	162	162

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 11.4 Response rates before school replacement, by adjudicated regions

		Weighted school participation rate before replacement (%) (SCHRRW1)	Weighted number of responding schools (weighted also by enrollment) (NUMW1)	Weighted number of schools sampled (responding + non-responding) (weighted also by enrollment) (DENW1)	Unweighted school participation rate before replacement (%) (SCHRRU1)	Number of responding schools (unweighted) (NUMU1)	Number of responding and non-responding schools (unweighted) (DENU1)
OECD	Belgium (Flemish community)	79.86	53531.06	67031.28	79.56	144	181
	Spain (Andalusia)	100.00	82758.93	82758.93	100.00	52	52
	Spain (Aragon)	100.00	11023.69	11023.69	100.00	51	51
	Spain (Asturias)	100.00	7014.26	7014.26	100.00	56	56
	Spain (Balearic Islands)	100.00	9155.92	9155.92	100.00	54	54
	Spain (Basque Country)	100.00	16853.96	16853.96	100.00	174	174
	Spain (Cantabria)	100.00	4474.20	4474.20	100.00	54	54
	Spain (Castile and Leon)	100.00	19736.92	19736.92	100.00	55	55
	Spain (Catalonia)	97.97	62481.18	63773.85	98.08	51	52
	Spain (Extremadura)	100.00	10977.66	10977.66	100.00	53	53
	Spain (Galicia)	100.00	19094.78	19094.78	100.00	56	56
	Spain (La Rioja)	100.00	2596.00	2596.00	100.00	54	54
	Spain (Madrid)	100.00	53017.19	53017.19	100.00	51	51
	Spain (Murcia)	99.28	14104.82	14206.83	98.11	52	53
	Spain (Navarre)	100.00	5478.28	5478.28	100.00	51	51
	United Kingdom (Scotland)	96.99	57927.71	59726.83	95.58	108	113
	United States (Connecticut)[1]	98.04	42460.41	43309.62	98.04	50	51
	United States (Florida)[1]	100.00	191020.11	191020.11	100.00	54	54
	United States (Massachusetts)[1]	100.00	69336.17	69336.17	100.00	49	49
Partners	Argentina (CABA)	93.91	33605.60	35783.59	94.12	48	51
	Russian Federation (Perm Territory)	100.00	28260.15	28260.15	100.00	63	63
	United Arab Emirates (Abu Dhabi)	99.54	16309.85	16384.98	99.16	118	119
	United Arab Emirates (Dubai)	98.51	13450.00	13654.00	97.16	205	211

1. Only public.

For calculating school response rates after replacement, the numerator consisted of all sampled schools (original plus replacement) with enrolled age-eligible students that participated (i.e. assessed a sample of PISA-eligible students and obtained a student response rate of at least 50%). The denominator consisted of all the schools in the numerator, plus those original sample schools that had age-eligible students enrolled, but that failed to assess at least 50% of PISA-eligible sample students and for which no replacement school participated. Schools that were included in the sampling frame, but were found to contain no age-eligible students, were omitted from the calculation of response rates. Replacement schools were included in rates only when they participated, and were replacing a refusing school that had age-eligible students.

In calculating weighted school response rates, each school received a weight equal to the product of its base weight (the reciprocal of its selection probability) and the number of age-eligible students enrolled in the school, as indicated on the school sampling frame.

With the use of Probability Proportional to Size sampling, where there are no certainty or small schools the product of the initial weight and the enrolment will be a constant, so in participating countries with few certainty school selections and no over-sampling or under-sampling of any explicit strata, weighted and unweighted rates are very similar. The weighted school response rate before replacement is given by the formula:

11.1

$$\begin{array}{c}\text{weighted school response rate} \\ \text{before replacement}\end{array} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i}$$

where Y denotes the set of responding original sample schools with age-eligible students, N denotes the set of eligible non-responding original sample schools, W_i denotes the base weight for school i, $W_i = 1/P_i$ where P_i denotes the school selection probability for school i, and E_i denotes the enrolment size of age-eligible students, as indicated on the sampling frame.

Table 11.5 Response rates after school replacement

		Weighted school participation rate after all replacement (%) (SCHRRW3)	Weighted number of responding schools (weighted also by enrollment) (NUMW3)	Weighted number of schools sampled (responding + non-responding) (weighted also by enrollment) (DENW3)	Unweighted school participation rate after all replacement (%) (SCHRRU3)	Number of responding schools (unweighted) (NUMU3)	Number of responding and non-responding schools (unweighted) (DENU3)
OECD	Australia	97.89	268630.57	274432.06	95.82	757	790
	Austria	100.00	88967.28	88967.28	100.00	191	191
	Belgium	96.64	115004.29	119005.58	95.92	282	294
	Canada	92.90	368599.91	396757.36	92.61	840	907
	Chile	98.83	236575.55	239370.42	98.66	221	224
	Czech Republic	99.61	88446.51	88796.51	99.33	295	297
	Denmark	95.51	67708.54	70892.45	92.62	339	366
	Estonia	100.00	12046.26	12046.26	100.00	206	206
	Finland	99.32	59911.57	60322.79	99.36	311	313
	France	96.58	703458.22	728400.97	96.54	223	231
	Germany	97.96	737778.25	753179.40	97.85	228	233
	Greece	98.86	100892.40	102053.43	97.92	188	192
	Hungary	99.45	101186.94	101751.21	98.08	204	208
	Iceland	99.34	4395.00	4424.00	95.00	133	140
	Ireland	99.32	57316.24	57711.25	98.92	183	185
	Israel	93.79	103074.94	109895.06	92.47	172	186
	Italy	97.37	522685.87	536820.92	96.27	1186	1232
	Japan	95.53	1123211.30	1175793.91	95.50	191	200
	Korea	99.86	661575.41	662509.92	99.36	156	157
	Luxembourg	100.00	5931.00	5931.00	100.00	42	42
	Mexico	95.31	1374615.03	1442234.18	93.98	1468	1562
	Netherlands	89.38	165634.94	185319.75	88.94	177	199
	New Zealand	89.33	52359.78	58615.69	89.85	177	197
	Norway	94.70	60270.13	63642.05	94.71	197	208
	Poland	97.95	393872.23	402116.47	96.81	182	188
	Portugal	95.83	122713.15	128050.31	95.90	187	195
	Slovak Republic	98.97	57599.34	58201.24	97.88	231	236
	Slovenia	98.12	18328.68	18680.14	94.90	335	353
	Spain	99.65	402604.32	403998.99	99.78	902	904
	Sweden	99.77	99535.54	99766.93	99.05	209	211
	Switzerland	98.33	82032.38	83423.80	97.16	410	422
	Turkey	99.94	944806.58	945356.78	99.41	169	170
	United Kingdom	89.23	624498.87	699838.56	91.82	505	550
	United States	77.21	3040660.50	3938076.78	77.78	161	207
Partners	Albania	100.00	49631.96	49631.96	100.00	204	204
	Argentina	95.86	580988.66	606069.17	95.63	219	229
	Brazil	95.44	2622293.26	2747687.51	94.47	837	886
	Bulgaria	99.81	57464.40	57574.24	99.47	187	188
	Colombia	97.44	596557.09	612261.21	96.97	352	363
	Costa Rica	98.94	64234.83	64919.87	98.96	191	193
	Croatia	99.94	45608.24	45636.39	99.39	163	164
	Cyprus[1, 2]	96.58	9485.00	9821.00	89.31	117	131
	Hong Kong-China	94.12	72064.20	76567.23	94.23	147	156
	Indonesia	98.01	2892364.78	2951027.68	98.10	206	210
	Jordan	100.00	119147.04	119147.04	100.00	233	233
	Kazakhstan	100.00	239767.02	239767.02	100.00	218	218
	Latvia	99.89	17428.27	17447.91	99.06	211	213
	Liechtenstein	100.00	382.00	382.00	100.00	12	12
	Lithuania	100.00	34604.37	34604.37	100.00	216	216
	Macao-China	100.00	5410.00	5410.00	100.00	45	45
	Malaysia	100.00	455543.42	455543.42	100.00	164	164
	Montenegro	100.00	8540.46	8540.46	100.00	51	51
	Peru	98.64	507601.62	514574.26	98.77	240	243
	Qatar	99.94	11333.00	11340.00	95.73	157	164
	Romania	100.00	139597.34	139597.34	100.00	178	178
	Russian Federation	100.00	1243563.91	1243563.91	100.00	227	227
	Serbia	95.44	69433.37	72752.34	95.00	152	160
	Shanghai-China	100.00	89831.82	89831.82	100.00	155	155
	Singapore	98.17	50945.18	51895.76	97.73	172	176
	Chinese Taipei	100.00	324667.48	324667.48	100.00	163	163
	Thailand	99.97	772452.27	772654.27	99.58	239	240
	Tunisia	99.30	129228.71	130141.41	99.35	152	153
	United Arab Emirates	99.40	46468.67	46747.80	98.48	453	460
	Uruguay	100.00	46009.37	46009.37	100.00	180	180
	Viet Nam	100.00	1068461.51	1068461.51	100.00	162	162

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 11.6 Response rates after school replacement, by adjudicated regions

		Weighted school participation rate after all replacement (%) (SCHRRW3)	Weighted number of responding schools (weighted also by enrollment) (NUMW3)	Weighted number of schools sampled (responding + non-responding) (weighted also by enrollment) (DENW3)	Unweighted school participation rate after all replacement (%) (SCHRRU3)	Number of responding schools (unweighted) (NUMU3)	Number of responding and non-responding schools (unweighted) (DENU3)
OECD	Belgium (Flemish community)	96.23	64491.57	67018.07	95.58	173	181
	Spain (Andalusia)	100.00	82758.93	82758.93	100.00	52	52
	Spain (Aragon)	100.00	11023.69	11023.69	100.00	51	51
	Spain (Asturias)	100.00	7014.26	7014.26	100.00	56	56
	Spain (Balearic Islands)	100.00	9155.92	9155.92	100.00	54	54
	Spain (Basque Country)	100.00	16853.96	16853.96	100.00	174	174
	Spain (Cantabria)	100.00	4474.20	4474.20	100.00	54	54
	Spain (Castile and Leon)	100.00	19736.92	19736.92	100.00	55	55
	Spain (Catalonia)	97.97	62481.18	63773.85	98.08	51	52
	Spain (Extremadura)	100.00	10977.66	10977.66	100.00	53	53
	Spain (Galicia)	100.00	19094.78	19094.78	100.00	56	56
	Spain (La Rioja)	100.00	2596.00	2596.00	100.00	54	54
	Spain (Madrid)	100.00	53017.19	53017.19	100.00	51	51
	Spain (Murcia)	99.28	14104.82	14206.83	98.11	52	53
	Spain (Navarre)	100.00	5478.28	5478.28	100.00	51	51
	United Kingdom (Scotland)	98.98	59135.50	59745.63	98.23	111	113
	United States (Connecticut)[1]	98.04	42460.41	43309.62	98.04	50	51
	United States (Florida)[1]	100.00	191020.11	191020.11	100.00	54	54
	United States (Massachusetts)[1]	100.00	69336.17	69336.17	100.00	49	49
Partners	Argentina (CABA)	93.91	33605.60	35783.59	94.12	48	51
	Russian Federation (Perm Territory)	100.00	28260.15	28260.15	100.00	63	63
	United Arab Emirates (Abu Dhabi)	99.54	16309.85	16384.98	99.16	118	119
	United Arab Emirates (Dubai)	98.51	13450.00	13654.00	97.16	205	211

1. Only public.

The weighted school response rate, after replacement, is given by the formula:

11.2

$$\begin{array}{c}\textit{weighted school response rate}\\ \textit{after replacement}\end{array} = \frac{\sum\limits_{i\in(Y\cup R)} W_iE_i}{\sum\limits_{i\in(Y\cup R\cup N)} W_iE_i}$$

where Y denotes the set of responding original sample schools, R denotes the set of responding replacement schools, for which the corresponding original sample school was eligible but was non-responding, N denotes the set of eligible refusing original sample schools, W_i denotes the base weight for school i, $W_i = 1/P_i$, where P_i denotes the school selection probability for school i, and for weighted rates, E_i denotes the enrolment size of age-eligible students, as indicated on the sampling frame.

For unweighted student response rates, the numerator is the number of students for whom assessment data were included in the results less those in schools with between 25 and 50% student participation. The denominator is the number of sampled students who were age-eligible, and not explicitly excluded as student exclusions. The exception is cases where participating countries applied different sampling rates across explicit strata. In these cases, unweighted rates were calculated in each stratum, and then weighted together according to the relative population size of 15-year-old students in each stratum.

For weighted student response rates, the same number of students appears in the numerator and denominator as for unweighted rates, but each student was weighted by its student base weight. This is given as the product of the school base weight – for the school in which the student was enrolled - and the reciprocal of the student selection probability within the school.

In countries with no over-sampling of any explicit strata, weighted and unweighted student participation rates are very similar.

Overall response rates are calculated as the product of school and student response rates. Although overall weighted and unweighted rates can be calculated, there is little value in presenting overall unweighted rates. The weighted rates indicate the proportion of the student population represented by the sample prior to making the school and student non-response adjustments.

Table 11.7 Response rates, students within schools after school replacement

		Weighted student participation rate after second replacement (%) (STURRW3)	Number of students assessed (Weighted) (NUMSTW3)	Number of students sampled (assessed + absent) (weighted) DENSTW3)	Unweighted student participation rate after second replacement (%) (STURRU3)	Number of students assessed (unweighted) (NUMSTU3)	Number of students sampled (assessed + absent) (unweighted) (DENSTU3)
OECD	Australia	86.78	213494.56	246011.73	84.10	17491	20799
	Austria	91.67	75393.38	82242.06	89.43	4756	5318
	Belgium	90.87	103914.08	114360.36	91.07	9649	10595
	Canada	80.76	261927.93	324328.31	81.26	20994	25835
	Chile	94.65	214558.03	226688.94	94.63	6857	7246
	Czech Republic	90.07	73536.20	81641.55	90.39	6528	7222
	Denmark	89.06	56095.69	62988.36	87.84	7463	8496
	Estonia	92.89	10806.70	11633.87	92.89	5867	6316
	Finland	90.73	54125.83	59652.83	90.19	8829	9789
	France	89.46	605371.31	676729.73	89.43	5641	6308
	Germany	93.24	692226.37	742415.62	93.18	4990	5355
	Greece	96.72	92443.95	95579.50	96.68	5125	5301
	Hungary	92.70	84032.16	90652.03	92.79	4810	5184
	Iceland	84.72	3503.00	4135.00	84.72	3503	4135
	Ireland	84.10	45114.70	53643.99	83.92	5016	5977
	Israel	90.02	91180.95	101288.39	90.10	6061	6727
	Italy	92.76	473104.40	510005.18	92.88	38084	41003
	Japan	96.10	1034802.88	1076786.01	96.10	6351	6609
	Korea	98.75	595460.85	603003.98	98.67	5033	5101
	Luxembourg	95.24	5260.00	5523.00	95.24	5260	5523
	Mexico	93.88	1193865.62	1271638.69	93.92	33786	35972
	Netherlands	84.97	148432.34	174697.50	85.02	4434	5215
	New Zealand	84.68	40396.62	47702.86	84.56	5248	6206
	Norway	90.88	51155.10	56285.99	90.88	4686	5156
	Poland	87.60	325389.10	371433.57	87.24	5629	6452
	Portugal	87.36	80718.62	92394.85	87.27	5608	6426
	Slovak Republic	93.75	50543.98	53912.39	93.96	5737	6106
	Slovenia	90.45	16145.66	17849.41	91.04	7211	7921
	Spain	89.88	334381.61	372041.80	91.10	26443	29027
	Sweden	92.17	87358.82	94784.50	92.18	4739	5141
	Switzerland	91.96	72116.05	78423.87	92.42	11218	12138
	Turkey	98.22	850830.05	866268.67	98.14	4847	4939
	United Kingdom	86.07	528231.00	613735.80	86.27	12638	14649
	United States	88.86	2429717.61	2734268.47	88.99	6094	6848
Partners	Albania	92.49	39275.20	42466.30	92.96	4743	5102
	Argentina	87.99	457293.79	519733.08	86.89	5804	6680
	Brazil	90.06	2133034.54	2368438.36	89.03	19877	22326
	Bulgaria	95.70	51819.24	54144.87	95.86	5280	5508
	Colombia	93.08	507177.93	544861.74	92.69	11164	12045
	Costa Rica	88.97	35525.05	39930.39	88.34	4582	5187
	Croatia	92.17	41912.35	45473.48	92.18	6153	6675
	Cyprus [1, 2]	93.30	8718.81	9344.44	93.04	5078	5458
	Hong Kong-China	93.09	62059.16	66664.63	93.11	4659	5004
	Indonesia	95.15	2478961.19	2605253.90	94.80	5579	5885
	Jordan	94.96	105493.14	111097.66	95.08	7038	7402
	Kazakhstan	98.87	206052.74	208411.39	98.88	5808	5874
	Latvia	90.90	14578.98	16038.62	91.20	5276	5785
	Liechtenstein	93.31	293.00	314.00	93.31	293	314
	Lithuania	92.09	30428.64	33041.96	92.03	4618	5018
	Macao-China	99.42	5335.00	5366.00	99.42	5335	5366
	Malaysia	93.96	405982.54	432079.74	94.00	5197	5529
	Montenegro	93.77	7233.40	7714.20	93.79	4799	5117
	Peru	96.01	398192.91	414727.82	95.93	6035	6291
	Qatar	99.73	10966.00	10996.00	99.73	10966	10996
	Romania	97.83	137859.65	140915.14	97.80	5074	5188
	Russian Federation	97.34	1141316.78	1172539.12	97.21	6418	6602
	Serbia	93.36	60365.69	64658.31	93.30	4681	5017
	Shanghai-China	98.46	83820.55	85127.32	98.56	6374	6467
	Singapore	94.31	47465.26	50329.51	94.21	5546	5887
	Chinese Taipei	96.33	281799.06	292542.40	96.29	6046	6279
	Thailand	98.90	695087.89	702817.85	98.88	6606	6681
	Tunisia	90.35	108342.15	119916.97	90.41	4391	4857
	United Arab Emirates	94.66	38227.62	40383.61	94.34	11460	12148
	Uruguay	90.02	35799.84	39770.88	90.02	5315	5904
	Viet Nam	99.86	955222.03	956517.48	99.86	4959	4966

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 11.8 Response rates, students within schools after school replacement, by adjudicated regions

		Weighted student participation rate after second replacement (%) (STURRW3)	Number of students assessed (weighted) (NUMSTW3)	Number of students sampled (assessed + absent) (weighted) (DENSTW3)	Unweighted student participation rate after second replacement (%) (STURRU3)	Number of students assessed (Unweighted) (NUMSTU3)	Number of students sampled (assessed + absent) (unweighted) (DENSTU3)
OECD	Belgium (Flemish community)	92.00	59239.12	64387.25	91.93	5967	6491
	Spain (Andalusia)	89.30	67139.79	75182.54	89.49	1635	1827
	Spain (Aragon)	91.80	9151.20	9968.62	92.06	1427	1550
	Spain (Asturias)	91.84	6548.97	7131.15	91.52	1640	1792
	Spain (Balearic Islands)	88.53	7410.38	8370.33	88.50	1455	1644
	Spain (Basque Country)	94.70	15278.68	16133.41	94.72	4805	5073
	Spain (Cantabria)	89.85	3895.28	4335.20	89.84	1547	1722
	Spain (Castile and Leon)	92.80	17116.82	18444.29	92.70	1639	1768
	Spain (Catalonia)	90.86	49812.54	54822.37	91.05	1587	1743
	Spain (Extremadura)	89.07	9276.51	10415.29	89.12	1573	1765
	Spain (Galicia)	92.34	16901.98	18304.01	92.28	1590	1723
	Spain (La Rioja)	90.34	2318.33	2566.31	91.19	1552	1702
	Spain (Madrid)	91.72	44838.19	48888.26	91.59	1677	1831
	Spain (Murcia)	85.50	11171.11	13065.51	85.36	1411	1653
	Spain (Navarre)	93.18	4889.81	5247.91	93.09	1562	1678
	United Kingdom (Scotland)	83.15	44330.11	53315.91	83.05	2945	3546
	United States (Connecticut)[1]	87.47	31751.08	36297.49	87.52	1697	1939
	United States (Florida)[1]	90.02	156920.38	174318.19	90.11	1896	2104
	United States (Massachusetts)[1]	90.04	51316.98	56990.91	90.30	1723	1908
Partners	Argentina (CABA)	84.85	26273.29	30964.89	84.41	1316	1559
	Russian Federation (Perm Territory)	96.86	26904.82	27776.21	96.71	1761	1821
	United Arab Emirates (Abu Dhabi)	94.75	14858.66	15681.80	94.47	3162	3347
	United Arab Emirates (Dubai)	92.92	9814.14	10562.20	93.08	4935	5302

1. Only public.

DESIGN EFFECTS AND EFFECTIVE SAMPLE SIZES

Surveys in education and especially international surveys rarely sample students by simply selecting a random sample of students (known as a simple random sample). Rather, a sampling design is used where schools are first selected and, within each selected school, classes or students are randomly sampled. Sometimes, geographic areas are first selected before sampling schools and students. This sampling design is usually referred to as a cluster sample or a multi-stage sample.

Selected students attending the same school cannot be considered as independent observations as assumed with a simple random sample because they are usually more similar to one another than to students attending other schools. For instance, the students are offered the same school resources, may have the same teachers and therefore are taught a common implemented curriculum, and so on. School differences are also larger if different educational programmes are not available in all schools. One expects to observe greater differences between a vocational school and an academic school than between two comprehensive schools.

Furthermore, it is well known that within a country, within sub-national entities and within a city, people tend to live in areas according to their financial resources. As children usually attend schools close to their home, it is likely that students attending the same school come from similar social and economic backgrounds.

A simple random sample of 4 000 students is thus likely to cover the diversity of the population better than a sample of 100 schools with 40 students observed within each school. It follows that the uncertainty associated with any population parameter estimate (i.e. standard error) will be larger for a clustered sample estimate than for a simple random sample estimate of the same size.

In the case of a simple random sample, the standard error of a mean estimate is equal to:

11.3

$$\sigma_{(\hat{\mu})} = \sqrt{\frac{\sigma^2}{n}}$$

For an infinite population of schools and infinite populations of students within schools, the standard error of a mean estimate from a cluster sample is equal to:

11.4

$$\sigma_{(\hat{\mu})} = \sqrt{\frac{\sigma^2_{schools}}{n_{schools}} + \frac{\sigma^2_{within}}{n_{schools}n_{students}}}$$

The standard error for the mean from a simple random sample is inversely proportional to the number of selected students. The standard error for the mean from a cluster sample is proportional to the variance that lies between clusters (i.e. schools) and within clusters and inversely proportional to the number of selected schools and the number of students selected per school.

It is usual to express the decomposition of the total variance into the between-school variance and the within-school variance by the coefficient of intraclass correlation, also denoted *Rho*. Mathematically, this index is equal to:

11.5

$$Rho = \frac{\sigma^2_{schools}}{\sigma^2_{schools} + \sigma^2_{within}}$$

This index provides an indication of the percentage of variance that lies between schools. A low intraclass correlation indicates that schools are performing similarly while higher values point towards large differences between school performance.

To limit the reduction of precision in the population parameter estimate, multi-stage sample designs usually use supplementary information to improve coverage of the population diversity. In PISA the following techniques were implemented to limit the increase in the standard error: *(i)* explicit and implicit stratification of the school sampling frame and *(ii)* selection of schools with probabilities proportional to their size. Complementary information generally cannot compensate totally for the increase in the standard error due to the multi-stage design however but will greatly reduce it.

Table 11.9 provides the standard errors on the PISA 2012 mathematics scale if the participating country sample was selected according to *(i)* a simple random sample; *(ii)* a multistage procedure without using complementary information (unstratified multistage sampling) and *(iii)* the balanced repeated replication (BRR) estimate for the actual PISA 2012 design, using Fay's method. It should be mentioned that the plausible value imputation variance was not included in these computations and thus only reflects sampling error.

Note that the values in Table 11.9 for the standard errors for the unstratified multistage design are overestimates for countries that had a school census (Cyprus,[1] Iceland, Liechtenstein, Luxembourg, Macao-China and Qatar) since these standard error estimates assume a sample of schools was collected. This also applies to the United Arab Emirates, as Dubai was a school census.

Also note that in some of the countries where the BRR estimates in Table 11.9 are greater than the values for the unstratified multistage sample, this is because of regional or other oversampling (Colombia [two regions], Mexico and Spain).

The BRR estimates in Table 11.9 are also greater than the values for the unstratified multistage sample for Canada, Denmark, Switzerland, the United Kingdom, Indonesia and Jordan. As described in the sampling design chapter, some countries have a substantial proportion of students attending schools that have fewer students than the target cluster size (*TCS*). Small schools were undersampled while schools in all large school strata were slightly oversampled. Of the countries mentioned, Indonesia has 20% of students, and Jordan 9%, in moderately small schools (with the number of students between half of TCS and TCS) and these schools are undersampled by a factor of 2. Demark also has 8.3% of students in this category of school, and a further 5.4% in very small schools (with the number of students less than half the TCS) which are undersampled by a factor of 4. Switzerland has 6.6% and 13.5% of students respectively in these school categories.

For the remaining two countries in Table 11.9 that have unbiased estimates, Canada and the United Kingdom, there are relatively few students in small schools. It is unclear in these cases why the BRR variance should be larger than the unstratified variance, though it is possible that the stratification undertaken possibly did not explain enough between-school variance in these countries.

It is usual to express the effect of the sampling design on the standard errors by a statistic referred to as the design effect. This corresponds to the ratio of the variance of the estimate obtained from the (more complex) sample to the variance

Table 11.9 Standard errors for the PISA 2012 mathematics scale

		Simple random sample	Unstratified multi-stage sample	BRR estimate for PISA sample
OECD	Australia	0.800	2.026	1.643
	Austria	1.341	5.348	2.668
	Belgium	1.103	4.674	2.145
	Canada	0.605	1.567	1.841
	Chile	0.975	4.585	3.067
	Czech Republic	1.301	4.492	2.852
	Denmark	0.949	2.245	2.295
	Estonia	1.170	2.729	2.021
	Finland	0.908	2.055	1.941
	France	1.435	5.067	2.455
	Germany	1.362	4.924	2.877
	Greece	1.226	4.297	2.502
	Hungary	1.350	5.521	3.193
	Iceland	1.553	3.416	1.702
	Ireland	1.194	2.972	2.248
	Israel	1.476	5.389	4.682
	Italy	0.526	2.084	2.025
	Japan	1.174	5.104	3.587
	Korea	1.397	5.081	4.579
	Luxembourg	1.316	8.505	1.095
	Mexico	0.404	1.266	1.353
	Netherlands	1.372	5.689	3.472
	New Zealand	1.521	4.022	2.205
	Norway	1.322	2.800	2.734
	Poland	1.332	3.681	3.617
	Portugal	1.242	4.032	3.812
	Slovak Republic	1.475	4.584	3.426
	Slovenia	1.192	3.986	1.233
	Spain	0.551	1.434	1.896
	Sweden	1.333	2.839	2.255
	Switzerland	0.890	2.878	3.041
	Turkey	1.308	5.625	4.830
	United Kingdom	0.840	2.308	3.296
	United States	1.274	3.701	3.597
Partners	Albania	1.329	2.049	2.002
	Argentina	0.999	3.864	3.530
	Brazil	0.564	1.955	1.937
	Bulgaria	1.292	5.227	3.988
	Colombia	0.780	2.671	2.887
	Costa Rica	1.008	3.181	3.042
	Croatia	1.250	4.604	3.539
	Cyprus[1, 2]	1.307	5.702	1.075
	Hong Kong-China	1.409	5.358	3.217
	Indonesia	0.952	3.442	4.036
	Jordan	0.925	3.046	3.116
	Kazakhstan	0.934	3.200	3.026
	Latvia	1.248	3.141	2.752
	Liechtenstein	5.576	18.445	3.952
	Lithuania	1.311	3.719	2.641
	Macao-China	1.294	9.048	0.965
	Malaysia	1.125	3.697	3.179
	Montenegro	1.200	6.881	1.054
	Peru	1.086	3.802	3.690
	Qatar	0.954	5.621	0.756
	Romania	1.142	4.405	3.761
	Russian Federation	1.194	3.292	3.035
	Serbia	1.325	5.121	3.389
	Shanghai-China	1.404	5.640	3.294
	Singapore	1.415	4.882	1.323
	Chinese Taipei	1.487	6.162	3.298
	Thailand	1.012	4.205	3.448
	Tunisia	1.178	4.528	3.915
	United Arab Emirates	0.835	3.070	2.426
	Uruguay	1.217	4.541	2.764
	Viet Nam	1.218	5.135	4.840

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

of the estimate that would be obtained from a simple random sample of the same number of sampling units. The design effect has two primary uses – in sample size estimation and in appraising the efficiency of more complex sampling plans (Cochran, 1977).

In PISA, as sampling variance has to be estimated by using the 80 BRR replicates, a design effect can be computed for a statistic *t* using:

11.6

$$Deff(t) = \frac{Var_{BRR}(t)}{Var_{SRS}(t)}$$

where $Var_{BRR}(t)$ is the sampling variance for the statistic *t* computed by the *BRR* replication method, and $Var_{SRS}(t)$ is the sampling variance for the same statistic *t* on the same data but considering the sample as a simple random sample.

Based on Table 11.9, the standard error on the mean estimate in mathematics in Australia is equal to 1.64 (rounded from 1.643). As the standard deviation of the mathematics performance is equal to 96.29, the design effect in Australia for the mean estimate in mathematics is therefore equal to:

11.7

$$Deff(t) = \frac{Var_{BRR}(t)}{Var_{SRS}(t)} = \frac{(1.64)^2}{\left[(96.29)^2/14\ 481\right]} = 4.20$$

The sampling variance on the mathematics performance mean in Australia is about four times larger than it would have been with a simple random sample of the same sample size. Note that the participating students are 14 485 as only this number were assessed for mathematics. The remaining students were assessed for financial literacy.

Another way to express the reduction of precision due to the complex sampling design is through the effective sample size, which expresses the simple random sample size that would give the same sampling variance as the one obtained from the actual complex sample design. The effective sample size for a statistic *t* is equal to:

11.8

$$Effn(t) = \frac{n}{Deff(t)} = \frac{n \times Var_{SRS}(t)}{Var_{Brr}(t)}$$

where n is equal to the actual number of units in the sample. The effective sample size in Australia for the mathematics performance mean is equal to:

11.9

$$Effn(t) = \frac{n}{Deff(t)} = \frac{n \times Var_{SRS}(t)}{Var_{Brr}(t)} = \frac{(96.29)^2}{(1.64)^2} = 3\ 447.3$$

In other words, a simple random sample of 3 447 students in Australia would have been as precise as the actual PISA 2012 sample for the national estimate of mean mathematics performance.

Variability of the design effect

Neither the design effect nor the effective sample size is a definitive characteristic of a sample. Both the design effect and the effective sample size vary with the variable and statistic of interest.

As previously stated, the sampling variance for estimates of the mean from a cluster sample is proportional to the intraclass correlation. In some countries, student performance varies between schools. Students in academic schools usually tend to perform well while on average student performance in vocational schools is lower. Let us now suppose that the height of the students was also measured. There are no reasons why students in academic schools should be taller than students in vocational schools, at least if there is no interaction between tracks and gender. For this particular variable, the expected value of the school variance should be equal to zero and therefore, the design effect should tend to one. As the segregation effect differs according to the variable, the design effect will also differ according to the variable.

The second factor that influences the size of the design effect is the choice of requested statistics. It tends to be large for means, proportions, and sums but substantially smaller for bivariate or multivariate statistics such as correlation and regression coefficients.

Design effects in PISA for performance variables

The notion of design effect as given earlier is extended and gives rise to five different design effect formulae to describe the influence of the sampling and test designs on the standard errors for statistics.

The total errors computed for the international PISA initial report *PISA 2012 Results* (OECD, 2014) that involves performance variables (plausible values) consist of two components: sampling variance and measurement variance. The standard error of proficiency estimates in PISA is inflated because the students were not sampled according to a simple random sample and also because the estimation of student proficiency includes some amount of measurement error.

For any statistic *r*, the population estimate and the sampling variance are computed for each plausible value and then combined as described in Chapter 9.

The five design effects and their respective effective sample sizes are defined as follows:

- Design Effect 1

11.10

$$Deff_1(r) = \frac{Var_{SRS}(r) + MVar(r)}{Var_{SRS}(r)}$$

where *MVar(r)* is the measurement error variance for the statistic *r*. This design effect shows the inflation of the total variance that would have occurred due to measurement error if in fact the samples were considered as a simple random sample. Table 11.10 provides, per domain and per cycle, the design effect 1 values, for any country that participated in at least one PISA cycle. Table 11.11 provides the corresponding effective sample size.

- Design Effect 2

11.11

$$Deff_2(r) = \frac{Var_{BRR}(r) + MVar(r)}{Var_{SRS}(r) + MVar(r)}$$

shows the inflation of the *total* variance due only to the use of a complex sampling design. Table 11.12 provides, for each domain and PISA cycle, the design effect 2 values, for each participating country. Table 11.13 provides the corresponding effective sample size.

- Design Effect 3

11.12

$$Deff_3(r) = \frac{Var_{BRR}(r)}{Var_{SRS}(r)}$$

shows the inflation of the sampling variance due to the use of a complex design. Table 11.14 provides, for each domain and PISA cycle, the design effect 3 values, for each participating country. Table 11.15 provides the corresponding effective sample size.

- Design Effect 4

11.13

$$Deff_4(r) = \frac{Var_{BRR}(r) + MVar(r)}{Var_{BRR}(r)}$$

shows the inflation of the total variance due to measurement error. Table 11.16 provides, for each domain and PISA cycle, the design effect 4 values, for each participating country. Table 11.17 provides the corresponding effective sample size.

[Part 1/2]

Table 11.10 Design effect 1 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1.30	1.49	1.20	1.22	1.11	1.14	1.16	1.10	1.12
	Austria	1.06	1.01	1.07	1.10	1.14	1.09	1.09	1.19	1.12
	Belgium	1.06	1.12	1.03	1.12	1.06	1.47	1.07	1.03	1.06
	Canada	1.09	1.12	1.10	1.49	1.51	1.82	1.30	1.08	1.13
	Chile	1.12	1.34	1.38				1.17	1.28	1.08
	Czech Republic	1.07	1.03	1.08	1.35	1.21	1.58	1.10	1.14	1.06
	Denmark	1.08	1.23	1.04	1.39	1.24	1.29	1.16	1.19	1.17
	Estonia							1.07	1.07	1.15
	Finland	1.14	1.25	1.24	1.16	1.25	1.28	1.12	1.60	1.23
	France	1.12	1.21	1.25	1.16	1.12	1.26	1.05	1.20	1.02
	Germany	1.13	1.06	1.22	1.05	1.01	1.12	1.07	1.14	1.08
	Greece	1.19	1.24	1.02	1.52	1.10	1.96	1.08	1.09	1.40
	Hungary	1.03	1.04	1.05	1.12	1.20	1.45	1.25	1.27	1.10
	Iceland	1.11	1.25	1.03	1.14	1.06	1.05	1.62	1.56	1.12
	Ireland	1.11	1.07	1.02	1.13	1.11	1.25	1.30	1.21	1.30
	Israel	1.47	1.15	1.33				1.12	1.23	1.04
	Italy	1.16	1.32	1.05	1.90	1.78	1.20	1.19	1.29	1.10
	Japan	1.11	1.10	1.17	1.31	1.09	1.10	1.17	1.03	1.05
	Korea	1.13	1.12	1.22	1.24	1.22	1.11	1.47	1.10	1.18
	Luxembourg	1.16	1.11	1.15	1.36	1.01	1.25	1.21	1.13	1.07
	Mexico	1.17	1.18	1.19	1.87	1.59	5.91	1.75	2.84	1.73
	Netherlands	1.06	1.08	1.02	1.29	1.09	1.29	1.36	1.19	1.18
	New Zealand	1.03	1.14	1.03	1.10	1.21	1.16	1.17	1.18	1.04
	Norway	1.06	1.24	1.06	1.26	1.03	1.14	1.10	1.13	1.06
	Poland	1.16	1.08	1.43	1.17	1.13	1.04	1.07	1.28	1.09
	Portugal	1.20	1.10	1.03	1.11	1.02	1.14	1.28	1.34	1.23
	Slovak Republic				1.03	1.14	1.02	1.13	1.43	1.13
	Slovenia							1.16	1.23	1.07
	Spain	1.17	1.03	1.04	1.83	1.36	1.38	1.33	2.18	1.92
	Sweden	1.20	1.12	1.13	1.17	1.06	1.43	1.65	1.06	1.10
	Switzerland	1.05	1.20	1.29	1.22	1.28	1.20	1.31	1.44	1.14
	Turkey				1.24	1.24	1.26	1.25	1.33	1.03
	United Kingdom	1.09	1.17	1.26	1.47	1.26	1.20	1.21	1.19	1.41
	United States	1.10	1.10	1.12	1.48	1.36	1.32		1.15	1.03
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	1.07	1.17	1.34						
	Argentina	1.18	1.17	1.31				1.29	1.33	1.11
	Azerbaijan							1.58	1.27	1.21
	Brazil	1.19	1.25	1.63	1.37	1.22	1.87	1.60	1.21	1.39
	Bulgaria	1.13	1.03	1.34				1.09	1.22	1.16
	Colombia							1.36	1.10	1.46
	Costa Rica									
	Croatia							1.17	1.12	1.12
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	1.05	1.10	1.12	1.07	1.42	1.19	1.09	1.13	1.03
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	1.48	1.24	1.29	1.98	1.46	1.70	1.29	1.94	1.16
	Jordan							1.51	1.20	1.07
	Kazakhstan									
	Kyrgyzstan							1.17	1.16	1.03
	Latvia	1.20	1.18	1.05	1.20	1.18	1.15	1.14	1.05	1.08
	Liechtenstein	1.10	1.15	1.04	1.05	1.21	1.16	1.10	1.22	1.13
	Lithuania							1.11	1.29	1.05
	Macao-China				1.29	1.05	1.19	1.21	1.39	1.09
	Macedonia	1.24	1.18	1.06						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							1.09	1.25	1.10
	Panama									
	Peru	1.10	1.19	2.02						
	Qatar							1.25	1.30	1.13
	Romania	1.25	1.14	1.15				1.40	1.39	1.07
	Russian Federation	1.16	1.15	1.14	1.22	1.28	1.15	1.42	1.23	1.08
	Russian Federation (Perm)									
	Serbia				1.11	1.29	1.36	1.14	1.33	1.05
	Shanghai-China									
	Singapore									
	Chinese Taipei							1.59	1.18	1.07
	Thailand	1.13	1.23	1.10	1.70	1.25	1.33	1.19	1.26	1.08
	Trinidad and Tobago									
	Tunisia				1.48	1.05	1.10	1.10	1.19	1.03
	United Arab Emirates									
	Uruguay				1.34	1.10	1.04	1.16	1.20	1.13
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.10 Design effect 1 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1.08	1.27	1.07	1.13	1.15	1.23
	Austria	1.14	1.12	1.08	1.14	1.06	1.23
	Belgium	1.15	1.22	1.26	1.17	1.20	1.16
	Canada	1.11	1.57	1.25	1.40	1.19	1.07
	Chile	1.29	1.14	1.14	1.19	1.26	1.10
	Czech Republic	1.23	1.11	1.09	1.40	1.35	1.63
	Denmark	1.11	1.09	1.32	1.09	1.05	1.03
	Estonia	1.21	1.16	1.27	1.08	1.19	1.11
	Finland	1.05	1.01	1.14	1.22	1.16	1.18
	France	1.04	1.10	1.05	1.09	1.28	1.06
	Germany	1.08	1.20	1.06	1.06	1.04	1.04
	Greece	1.31	1.21	1.60	1.10	1.15	1.14
	Hungary	1.00	1.07	1.05	1.16	1.21	1.14
	Iceland	1.03	1.13	1.03	1.13	1.12	1.52
	Ireland	1.02	1.02	1.15	1.11	1.10	1.21
	Israel	1.26	1.06	1.29	1.21	1.08	1.20
	Italy	1.23	1.21	1.52	1.20	1.25	1.53
	Japan	1.06	1.09	1.11	1.06	1.03	1.23
	Korea	1.27	1.06	1.45	1.13	1.12	1.10
	Luxembourg	1.22	1.23	1.21	1.68	1.29	1.39
	Mexico	1.39	1.03	1.68	1.19	1.11	1.10
	Netherlands	1.14	1.07	1.21	1.02	1.26	1.05
	New Zealand	1.09	1.10	1.05	1.10	1.14	1.02
	Norway	1.20	1.13	1.21	1.31	1.04	1.04
	Poland	1.12	1.21	1.30	1.11	1.02	1.22
	Portugal	1.06	1.16	1.17	1.10	1.19	1.21
	Slovak Republic	1.10	1.03	1.10	1.21	1.12	1.15
	Slovenia	1.08	1.19	1.16	1.34	1.25	1.36
	Spain	1.10	1.68	1.29	1.25	1.43	1.17
	Sweden	1.08	1.16	1.05	1.03	1.05	1.09
	Switzerland	1.15	1.81	1.28	1.16	1.22	1.36
	Turkey	1.16	1.07	1.15	1.09	1.17	1.20
	United Kingdom	1.09	1.20	1.19	1.22	1.27	1.21
	United States	1.10	1.07	1.11	1.11	1.39	1.09
	United States (Connecticut)[1]				1.26	1.05	1.32
	United States (Florida)[1]				1.03	1.08	1.04
	United States (Massachusetts)[1]				1.17	1.04	1.08
Partners	Albania	1.03	1.38	1.14	1.25	1.26	1.06
	Argentina	1.09	1.06	1.03	1.10	1.06	1.14
	Azerbaijan	1.35	1.14	1.39			
	Brazil	1.11	1.31	1.34	1.05	1.07	1.46
	Bulgaria	1.20	1.04	1.06	1.09	1.06	1.06
	Colombia	1.11	1.49	1.26	1.20	1.05	1.27
	Costa Rica	1.14	1.18	1.26	1.65	1.20	1.54
	Croatia	1.04	1.18	1.14	1.14	1.02	1.06
	Cyprus[2, 3]				1.04	1.04	1.10
	Georgia	1.23	1.10	1.12			
	Hong Kong-China	1.08	1.05	1.01	1.17	1.14	1.09
	India (Himachal Pradesh)	1.20	1.17	1.01			
	India (Tamil Nadu)	1.09	1.33	1.15			
	Indonesia	1.24	1.21	1.46	1.29	1.37	2.12
	Jordan	1.04	1.09	1.14	1.13	1.05	1.11
	Kazakhstan	1.15	1.25	1.09	1.38	1.24	1.45
	Kyrgyzstan	1.04	1.08	1.19			
	Latvia	1.19	1.08	1.40	1.07	1.11	1.20
	Liechtenstein	1.04	1.14	1.07	1.17	1.09	1.08
	Lithuania	1.09	1.10	1.12	1.04	1.11	1.16
	Macao-China	1.24	1.08	1.45	1.09	1.01	1.05
	Macedonia						
	Malaysia	1.05	1.18	1.26	1.19	1.26	1.29
	Malta	1.05	1.05	1.19			
	Mauritius	1.11	1.12	1.18			
	Moldova	1.25	1.20	1.18			
	Montenegro	1.10	1.21	1.31	1.18	1.16	1.12
	Panama	1.44	1.18	1.07			
	Peru	1.14	2.01	1.86	1.23	1.09	1.37
	Qatar	1.01	1.05	1.25	1.09	1.10	1.02
	Romania	1.01	1.31	1.09	1.20	1.43	1.08
	Russian Federation	1.15	1.06	1.22	1.21	1.09	1.12
	Russian Federation (Perm)				1.13	1.15	1.25
	Serbia	1.13	1.03	1.04	1.19	1.08	1.19
	Shanghai-China	1.13	1.06	1.21	1.54	1.21	1.45
	Singapore	1.07	1.41	1.24	1.06	1.05	1.10
	Chinese Taipei	1.13	1.04	1.17	1.18	1.12	1.10
	Thailand	1.14	1.02	1.28	1.21	1.11	1.23
	Trinidad and Tobago	1.02	1.35	1.14			
	Tunisia	1.08	1.10	1.10	1.20	1.08	1.21
	United Arab Emirates	1.23	1.23	1.48	1.15	1.23	1.07
	Uruguay	1.13	1.38	1.43	1.11	1.06	1.04
	Venezuela (Miranda)	1.05	1.07	1.18			
	Viet Nam				1.05	1.04	1.08

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.11 Effective sample size 1 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	3982.98	1922.54	2373.94	10327.97	11334.89	11054.93	12175.93	12840.75	12653.94
	Austria	4483.16	2620.10	2499.64	4195.29	4040.39	4210.54	4507.65	4140.59	4398.95
	Belgium	6302.33	3366.19	3613.01	7860.91	8290.72	5986.82	8256.26	8613.58	8363.93
	Canada	27294.23	14681.89	15047.11	18723.40	18559.00	15319.62	17464.56	21010.67	20047.99
	Chile	4371.55	2026.64	1959.44				4489.59	4085.75	4854.72
	Czech Republic	5018.59	2964.40	2841.42	4680.72	5221.12	4006.13	5376.93	5194.96	5603.99
	Denmark	3923.55	1935.63	2255.75	3031.77	3402.08	3258.98	3892.26	3810.44	3877.16
	Estonia							4527.59	4554.38	4248.12
	Finland	4270.16	2163.08	2179.57	5009.27	4626.51	4536.95	4202.73	2940.60	3835.58
	France	4189.42	2153.31	2080.42	3707.19	3850.71	3404.12	4470.35	3923.30	4616.93
	Germany	4473.45	2681.79	2340.72	4453.86	4603.09	4155.56	4566.28	4290.17	4515.17
	Greece	3929.58	2108.13	2552.96	3053.97	4191.93	2365.63	4497.01	4458.84	3484.71
	Hungary	4743.02	2701.20	2678.39	4271.51	3977.70	3277.53	3602.75	3543.41	4089.22
	Iceland	3044.83	1505.27	1804.12	2940.36	3163.72	3178.72	2340.60	2421.28	3386.89
	Ireland	3473.71	1984.38	2097.43	3433.59	3482.92	3096.49	3528.03	3803.64	3530.23
	Israel	3063.18	2161.42	1884.00				4076.52	3738.80	4390.02
	Italy	4279.64	2101.47	2629.13	6122.91	6555.34	9668.31	18288.26	16892.07	19776.44
	Japan	4752.55	2655.44	2489.04	3595.31	4307.76	4296.25	5086.10	5773.51	5679.70
	Korea	4412.97	2469.80	2263.79	4379.00	4456.87	4897.91	3518.97	4706.33	4388.00
	Luxembourg	3043.06	1760.61	1698.16	2889.77	3871.84	3134.89	3782.54	4032.48	4282.74
	Mexico	3945.33	2181.29	2149.30	15997.66	18839.24	5073.84	17695.77	10893.95	17860.73
	Netherlands	2368.53	1279.70	1364.29	3102.67	3676.16	3093.18	3583.17	4106.08	4142.42
	New Zealand	3548.97	1793.29	1974.15	4101.85	3741.98	3891.50	4122.38	4073.13	4628.65
	Norway	3895.23	1857.37	2181.11	3214.90	3946.20	3569.96	4252.80	4152.82	4439.13
	Poland	3157.62	1822.78	1424.55	3747.88	3894.01	4221.95	5166.51	4344.48	5105.06
	Portugal	3836.42	2322.65	2470.80	4166.25	4533.92	4051.84	4005.22	3802.79	4153.20
	Slovak Republic				7111.33	6466.20	7182.62	4182.98	3305.75	4193.52
	Slovenia							5693.01	5372.55	6145.89
	Spain	5323.48	3330.42	3339.28	5898.67	7917.90	7805.60	14768.23	9004.61	10226.15
	Sweden	3668.68	2206.52	2163.14	3960.37	4361.79	3239.77	2690.43	4179.79	4043.95
	Switzerland	5798.25	2841.45	2625.89	6883.34	6595.95	7032.56	9334.98	8456.35	10732.37
	Turkey				3901.43	3904.83	3863.79	3958.90	3728.54	4788.64
	United Kingdom	8551.97	4450.45	4098.53	6489.42	7588.24	7964.35	10845.28	11047.20	9296.56
	United States	3499.74	1949.72	1894.21	3681.91	4014.60	4139.33		4899.49	5426.07
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	4653.33	2379.20	2062.73						
	Argentina	3363.43	1900.68	1686.46				3355.42	3258.34	3895.54
	Azerbaijan							3277.75	4074.61	4288.49
	Brazil	4112.08	2175.18	1660.08	3244.15	3638.84	2381.43	5804.48	7668.39	6671.51
	Bulgaria	4128.43	2532.96	1897.44				4113.95	3688.25	3873.06
	Colombia							3304.74	4053.93	3073.88
	Costa Rica									
	Croatia							4438.25	4658.71	4666.41
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	4198.67	2222.91	2180.87	4170.70	3162.41	3776.99	4280.73	4107.61	4488.07
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	4979.59	3303.86	3153.42	5435.85	7375.21	6339.51	8244.32	5499.96	9190.51
	Jordan							4319.24	5434.01	6066.34
	Kazakhstan									
	Kyrgyzstan							5031.19	5094.87	5705.67
	Latvia	3240.09	1825.70	2058.50	3850.50	3919.78	4026.24	4135.75	4481.34	4367.51
	Liechtenstein	286.31	152.59	169.77	315.91	274.49	285.28	308.81	277.81	299.80
	Lithuania							4255.13	3675.26	4534.91
	Macao-China				970.50	1189.20	1052.93	3943.84	3423.77	4377.34
	Macedonia	3629.17	2149.35	2386.79						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							4101.60	3570.42	4038.81
	Panama									
	Peru	4019.66	2067.32	1218.17						
	Qatar							5030.18	4813.54	5548.37
	Romania	3863.43	2351.23	2348.84				3667.73	3681.21	4805.00
	Russian Federation	5770.68	3231.76	3252.25	4888.13	4666.97	5177.71	4090.65	4710.84	5354.36
	Russian Federation (Perm)									
	Serbia				3976.96	3423.75	3246.56	4215.59	3617.39	4578.02
	Shanghai-China									
	Singapore									
	Chinese Taipei							5534.93	7447.72	8270.44
	Thailand	4726.39	2406.19	2698.31	3072.75	4177.30	3933.88	5193.41	4897.57	5720.87
	Trinidad and Tobago									
	Tunisia				3181.44	4496.54	4283.93	4225.00	3889.90	4525.95
	United Arab Emirates									
	Uruguay				4344.33	5307.73	5608.11	4174.89	4048.89	4293.24
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.11 Effective sample size 1 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	13247.83	11263.38	13350.55	12834.52	12636.69	11739.03
	Austria	5781.28	5864.68	6105.52	4179.47	4495.10	3855.92
	Belgium	7369.19	6950.49	6762.98	7368.67	7156.42	7439.97
	Canada	20822.96	14747.39	18578.50	15432.01	18160.05	20102.32
	Chile	4395.77	4971.65	4965.65	5741.31	5433.92	6206.47
	Czech Republic	4926.96	5463.52	5539.22	3812.94	3955.16	3270.54
	Denmark	5317.55	5429.77	4499.16	6878.61	7117.40	7232.52
	Estonia	3897.26	4063.20	3717.90	4431.13	4017.61	4296.41
	Finland	5531.29	5748.44	5106.82	7213.81	7582.30	7512.72
	France	4145.05	3923.93	4078.36	4251.03	3598.59	4340.07
	Germany	4624.06	4141.17	4680.81	4733.03	4807.21	4812.39
	Greece	3785.65	4114.65	3103.50	4662.26	4475.59	4515.41
	Hungary	4589.00	4303.43	4386.23	4131.88	3981.69	4233.54
	Iceland	3528.25	3226.41	3555.31	3099.52	3130.57	2313.92
	Ireland	3860.28	3842.11	3410.50	4511.40	4557.90	4154.70
	Israel	4577.91	5421.98	4474.80	4162.70	4670.30	4214.37
	Italy	25079.77	25573.44	20375.65	25977.37	24762.39	20309.74
	Japan	5743.68	5607.14	5483.71	5981.32	6187.75	5165.53
	Korea	3937.75	4727.31	3449.26	4435.24	4512.79	4575.57
	Luxembourg	3783.09	3768.13	3829.72	3137.69	4078.96	3782.47
	Mexico	27507.37	37285.34	22716.84	28366.52	30470.79	30614.77
	Netherlands	4164.04	4439.43	3936.33	4374.84	3535.72	4245.36
	New Zealand	4275.50	4207.35	4407.92	3914.00	3766.97	4196.96
	Norway	3867.55	4141.63	3849.67	3567.18	4509.74	4519.26
	Poland	4394.30	4067.09	3794.56	4146.79	4505.85	3775.59
	Portugal	5930.90	5445.76	5394.61	5200.71	4801.44	4742.54
	Slovak Republic	4157.80	4416.24	4130.28	3850.74	4164.42	4078.31
	Slovenia	5717.21	5163.72	5300.31	4412.31	4724.39	4356.01
	Spain	23562.15	15372.35	20137.93	20275.35	17697.73	21585.32
	Sweden	4246.69	3938.55	4335.50	4578.18	4493.33	4349.38
	Switzerland	10272.74	6535.94	9251.39	9714.83	9169.94	8257.80
	Turkey	4314.78	4679.53	4350.58	4450.25	4137.93	4049.32
	United Kingdom	11179.46	10187.16	10240.64	10404.72	9980.53	10496.71
	United States	4765.05	4902.46	4695.95	4500.33	3586.12	4554.65
	United States (Connecticut)[1]				1348.28	1608.86	1289.67
	United States (Florida)[1]				1843.65	1751.88	1818.18
	United States (Massachusetts)[1]				1475.61	1656.90	1595.29
Partners	Albania	4453.28	3336.06	4043.23	3793.03	3752.18	4483.38
	Argentina	4368.09	4504.99	4635.88	5372.79	5593.69	5198.33
	Azerbaijan	3483.42	4109.27	3378.06			
	Brazil	18196.94	15308.34	14969.78	18338.34	17864.35	13196.05
	Bulgaria	3760.51	4344.15	4269.00	4835.88	4989.54	4990.44
	Colombia	7142.48	5333.68	6309.13	7563.75	8634.03	7169.30
	Costa Rica	4013.76	3867.46	3637.21	2791.69	3841.51	2982.95
	Croatia	4806.84	4228.21	4387.22	4383.84	4924.28	4727.88
	Cyprus[2, 3]				4902.23	4889.36	4606.05
	Georgia	3765.28	4218.72	4156.03			
	Hong Kong-China	4473.61	4597.67	4778.50	4004.48	4086.49	4268.16
	India (Himachal Pradesh)	1345.09	1383.57	1602.74			
	India (Tamil Nadu)	2940.86	2418.75	2789.94			
	Indonesia	4135.38	4249.07	3518.28	4345.24	4102.70	2646.99
	Jordan	6261.09	5951.09	5665.51	6230.44	6728.56	6348.08
	Kazakhstan	4701.84	4313.77	4969.56	4203.48	4684.76	4006.25
	Kyrgyzstan	4790.79	4613.12	4195.14			
	Latvia	3769.88	4172.13	3204.88	4013.83	3876.59	3602.12
	Liechtenstein	315.08	288.56	307.09	249.75	269.56	270.98
	Lithuania	4151.20	4126.90	4041.34	4459.36	4173.79	3976.47
	Macao-China	4804.07	5505.51	4098.97	4880.58	5261.37	5076.53
	Macedonia						
	Malaysia	4765.88	4226.06	3965.60	4352.06	4134.28	4014.78
	Malta	3300.13	3276.22	2895.28			
	Mauritius	4196.87	4167.35	3930.94			
	Moldova	4155.83	4330.01	4408.55			
	Montenegro	4398.79	3986.15	3679.98	4024.07	4076.43	4246.58
	Panama	2748.42	3370.38	3698.24			
	Peru	5263.01	2971.64	3223.23	4893.66	5533.98	4392.96
	Qatar	9032.59	8612.18	7285.23	10029.02	9982.51	10796.67
	Romania	4721.70	3642.12	4388.03	4234.28	3559.68	4681.95
	Russian Federation	4629.00	4997.11	4355.84	4329.33	4803.58	4688.86
	Russian Federation (Perm)				1565.25	1527.19	1412.64
	Serbia	4870.95	5373.29	5317.12	3949.60	4342.91	3928.20
	Shanghai-China	4524.98	4845.08	4233.83	3353.65	4281.13	3563.79
	Singapore	4923.70	3749.40	4277.07	5242.12	5277.56	5019.59
	Chinese Taipei	5157.00	5581.21	4971.33	5112.88	5404.91	5493.79
	Thailand	5445.50	6098.22	4880.55	5464.48	5961.17	5370.66
	Trinidad and Tobago	4688.38	3548.06	4199.46			
	Tunisia	4572.80	4493.97	4498.59	3671.73	4066.53	3654.59
	United Arab Emirates	8856.32	8814.91	7346.32	9978.71	9384.76	10736.49
	Uruguay	5270.57	4326.17	4160.21	4781.04	5019.15	5098.47
	Venezuela (Miranda)	2758.22	2719.86	2464.79			
	Viet Nam				4729.04	4788.69	4580.97

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.12 Design effect 2 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	4.77	2.89	3.22	4.92	5.75	4.69	5.89	8.32	6.44
	Austria	2.98	1.93	1.95	5.58	4.97	5.29	6.41	6.01	7.08
	Belgium	6.96	4.54	5.39	4.33	3.59	3.18	6.31	6.68	5.20
	Canada	7.41	4.05	4.70	7.29	8.08	6.34	11.21	11.04	9.33
	Chile	6.96	3.13	2.59				10.50	11.22	10.77
	Czech Republic	3.04	2.46	1.90	6.15	7.13	4.51	7.59	6.15	6.99
	Denmark	2.26	1.53	1.67	3.09	3.07	2.78	4.93	3.63	4.32
	Estonia							5.37	5.31	3.86
	Finland	3.55	1.54	1.80	2.06	2.30	2.04	2.94	2.37	2.13
	France	3.70	1.99	2.01	2.83	2.87	2.48	6.83	4.32	5.05
	Germany	2.20	1.62	1.33	4.29	4.81	4.42	7.09	6.54	6.51
	Greece	10.29	5.60	6.51	4.70	7.24	3.41	6.98	4.61	4.28
	Hungary	8.41	4.53	4.42	3.08	3.66	2.66	4.36	3.56	3.77
	Iceland	0.75	1.06	1.10	0.74	0.78	0.75	0.94	1.02	0.97
	Ireland	4.16	2.09	2.52	3.16	2.87	2.59	5.16	4.38	4.02
	Israel	18.44	10.96	9.86				6.00	6.12	4.85
	Italy	4.35	2.21	2.54	5.59	6.77	8.14	9.10	9.59	8.83
	Japan	17.53	10.60	9.12	4.97	6.87	6.16	6.46	7.78	6.45
	Korea	5.33	2.65	2.52	6.14	5.47	6.07	6.56	7.77	6.10
	Luxembourg	0.77	0.81	0.98	0.64	0.43	0.67	0.62	0.53	0.51
	Mexico	5.88	3.60	3.66	29.59	34.24	8.22	18.09	12.83	20.21
	Netherlands	3.39	2.17	2.32	3.51	4.21	3.15	3.28	3.50	3.40
	New Zealand	2.35	1.82	1.12	2.27	1.97	2.00	3.33	2.67	2.92
	Norway	2.85	1.70	1.81	2.36	2.63	2.74	3.89	3.45	4.65
	Poland	6.29	5.20	3.99	3.37	3.00	3.30	4.02	3.46	3.47
	Portugal	8.30	4.63	4.98	6.75	6.84	5.56	5.20	4.35	4.84
	Slovak Republic				8.09	8.32	9.47	3.54	2.95	3.23
	Slovenia							0.71	0.73	0.79
	Spain	5.44	3.96	3.19	4.38	5.87	5.31	9.34	6.21	8.21
	Sweden	2.10	1.53	1.57	2.54	3.18	2.11	3.29	3.01	2.57
	Switzerland	10.04	5.49	5.18	8.23	7.80	8.26	9.88	8.86	10.88
	Turkey				14.39	16.15	14.55	8.11	10.30	10.19
	United Kingdom	5.55	3.31	3.07	4.46	5.25	4.81	5.31	6.41	4.27
	United States	15.82	11.77	9.91	3.73	3.85	3.80		9.83	8.61
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	5.10	1.97	1.94						
	Argentina	27.72	11.50	10.32				11.18	12.41	14.05
	Azerbaijan							6.48	9.03	10.49
	Brazil	5.32	3.14	2.16	5.49	8.54	4.65	7.75	7.79	6.50
	Bulgaria	9.54	6.78	4.35				14.20	13.56	12.70
	Colombia							7.34	7.48	4.87
	Costa Rica									
	Croatia							4.43	3.75	3.79
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	5.10	2.69	2.73	7.88	6.48	7.74	3.75	3.36	3.27
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	15.08	9.47	8.71	10.69	17.38	14.12	51.68	27.19	61.43
	Jordan							5.21	8.47	6.05
	Kazakhstan									
	Kyrgyzstan							5.83	7.83	6.98
	Latvia	8.62	3.40	6.80	6.34	6.90	7.08	6.99	5.99	5.42
	Liechtenstein	0.52	0.81	0.95	0.50	0.47	0.50	0.52	0.57	0.54
	Lithuania							4.15	3.90	4.25
	Macao-China				1.01	1.31	1.25	0.81	0.82	0.80
	Macedonia	1.55	1.60	1.53						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							0.75	0.92	0.72
	Panama									
	Peru	8.47	3.46	2.41						
	Qatar							0.61	0.61	0.58
	Romania	4.45	3.20	2.98				9.57	9.25	12.87
	Russian Federation	11.79	8.90	7.42	8.70	9.66	8.92	8.80	8.79	8.97
	Russian Federation (Perm)									
	Serbia				7.59	6.73	5.80	6.00	5.30	5.82
	Shanghai-China									
	Singapore									
	Chinese Taipei							8.86	11.79	11.80
	Thailand	8.44	4.57	4.27	3.97	5.59	4.34	5.21	4.03	4.41
	Trinidad and Tobago									
	Tunisia				2.74	4.30	3.68	7.21	7.21	5.83
	United Arab Emirates									
	Uruguay				3.47	5.76	3.95	3.35	2.79	3.64
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 11.12 Design effect 2 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	7.40	8.15	8.30	3.38	3.68	3.60
	Austria	5.01	4.49	6.19	3.78	3.74	3.31
	Belgium	3.93	3.26	3.90	3.59	3.15	3.46
	Canada	5.61	5.01	6.05	6.74	7.80	9.05
	Chile	6.28	7.25	6.39	7.96	7.84	7.89
	Czech Republic	4.84	5.04	5.17	3.99	3.57	3.50
	Denmark	3.27	4.86	3.28	6.58	5.56	6.33
	Estonia	3.90	4.07	3.73	2.83	2.51	2.54
	Finland	3.76	3.97	3.50	4.57	3.92	4.20
	France	4.41	3.69	5.02	2.85	2.28	2.88
	Germany	3.63	3.51	3.62	4.49	4.29	4.66
	Greece	7.81	7.74	6.04	5.10	3.63	5.63
	Hungary	5.69	6.04	5.79	4.88	4.63	4.53
	Iceland	0.76	0.75	0.78	1.05	1.07	1.05
	Ireland	3.77	3.39	3.86	3.95	3.22	2.99
	Israel	4.86	5.38	3.81	7.97	9.30	8.92
	Italy	6.74	10.17	6.86	10.71	11.80	8.80
	Japan	6.85	7.01	6.42	8.28	9.10	7.35
	Korea	7.52	9.57	6.03	9.20	9.64	9.12
	Luxembourg	0.55	0.55	0.53	0.68	0.54	0.60
	Mexico	14.66	19.95	12.17	10.08	10.11	10.44
	Netherlands	14.05	12.60	12.50	6.08	5.08	5.77
	New Zealand	2.24	2.43	2.54	2.02	1.85	1.75
	Norway	3.10	3.27	3.24	3.65	4.11	4.33
	Poland	3.75	4.20	2.93	5.35	7.22	4.94
	Portugal	7.40	5.54	6.51	8.37	7.91	8.45
	Slovak Republic	3.31	4.54	4.04	6.16	4.81	5.18
	Slovenia	0.74	0.85	0.79	0.77	0.85	0.88
	Spain	12.56	8.32	11.09	8.72	8.26	9.69
	Sweden	3.62	3.76	3.22	3.62	2.71	3.95
	Switzerland	7.02	7.22	7.97	7.87	9.53	7.35
	Turkey	7.97	10.58	8.64	10.71	11.63	9.63
	United Kingdom	6.39	7.85	6.65	13.49	12.14	12.06
	United States	6.81	7.59	6.54	7.45	5.74	7.36
	United States (Connecticut)[1]				5.78	6.28	4.38
	United States (Florida)[1]				8.83	8.15	8.75
	United States (Massachusetts)[1]				5.82	6.70	6.10
Partners	Albania	7.28	6.41	7.95	2.89	1.80	2.75
	Argentina	8.00	8.63	9.32	7.98	11.83	10.59
	Azerbaijan	6.77	7.66	5.75			
	Brazil	15.32	13.24	12.55	10.12	10.95	8.88
	Bulgaria	13.09	15.22	13.15	12.48	9.00	10.92
	Colombia	13.34	9.85	12.60	12.84	13.02	11.46
	Costa Rica	6.35	6.68	5.00	6.19	7.60	5.18
	Croatia	5.16	5.19	4.90	6.46	7.88	6.22
	Cyprus[2, 3]				0.56	0.65	0.69
	Georgia	3.28	4.49	4.32			
	Hong Kong-China	2.84	3.77	4.74	4.28	4.56	4.17
	India (Himachal Pradesh)	3.65	4.76	5.19			
	India (Tamil Nadu)	14.43	13.26	12.08			
	Indonesia	13.06	11.90	10.64	13.52	13.12	8.27
	Jordan	8.31	11.99	8.93	9.43	10.86	9.03
	Kazakhstan	5.36	5.72	6.50	5.58	8.46	6.45
	Kyrgyzstan	5.00	5.83	4.36			
	Latvia	5.15	6.26	4.95	3.17	4.38	4.40
	Liechtenstein	0.36	0.62	0.47	0.54	0.46	0.46
	Lithuania	3.18	3.64	4.79	3.67	3.67	3.53
	Macao-China	0.66	0.63	0.75	0.59	0.55	0.59
	Macedonia						
	Malaysia	5.99	5.79	4.95	6.90	6.35	5.85
	Malta	0.58	0.62	0.64			
	Mauritius	0.52	0.50	0.58			
	Moldova	4.18	5.81	5.35			
	Montenegro	1.50	2.28	1.99	0.66	0.66	0.67
	Panama	11.92	14.28	15.02			
	Peru	8.50	5.87	4.93	10.55	10.59	9.19
	Qatar	0.40	0.44	0.54	0.53	0.57	0.53
	Romania	9.76	6.74	7.97	8.22	7.61	7.98
	Russian Federation	6.40	7.51	5.84	4.65	5.93	5.30
	Russian Federation (Perm)				6.38	5.80	5.50
	Serbia	4.11	5.58	4.21	5.45	6.06	5.98
	Shanghai-China	4.04	3.63	3.37	4.27	4.56	4.88
	Singapore	0.58	0.71	0.73	0.97	0.83	1.05
	Chinese Taipei	4.67	5.84	4.59	5.62	4.40	4.33
	Thailand	7.35	10.16	6.83	8.51	10.49	7.90
	Trinidad and Tobago	0.56	0.59	0.55			
	Tunisia	5.23	6.66	4.90	9.64	10.19	7.04
	United Arab Emirates	7.36	6.29	5.59	6.85	6.89	9.62
	Uruguay	3.63	3.47	2.96	5.19	4.87	4.29
	Venezuela (Miranda)	7.57	7.00	6.26			
	Viet Nam				16.65	15.25	14.23

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]
Table 11.13 Effective sample size 2 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1085.02	990.51	888.58	2548.91	2184.41	2674.75	2405.61	1702.55	2201.02
	Austria	1589.97	1370.28	1370.49	824.15	924.59	868.46	769.07	819.95	696.30
	Belgium	957.94	834.07	690.15	2030.99	2451.88	2766.80	1403.53	1326.24	1704.76
	Canada	4008.64	4071.63	3505.74	3833.69	3458.14	4407.16	2020.47	2051.63	2428.29
	Chile	702.49	870.35	1046.81				498.32	466.53	486.02
	Czech Republic	1765.84	1245.55	1611.27	1027.05	886.58	1400.09	781.33	964.10	848.27
	Denmark	1875.33	1555.71	1405.18	1366.62	1373.55	1519.91	918.72	1249.40	1049.48
	Estonia							906.67	916.90	1258.92
	Finland	1370.18	1751.35	1509.58	2819.73	2519.26	2844.06	1605.60	1991.14	2213.08
	France	1261.90	1305.32	1290.45	1522.12	1498.43	1733.40	690.21	1092.50	933.70
	Germany	2308.74	1747.14	2142.45	1086.88	968.52	1053.24	690.25	748.08	751.76
	Greece	454.23	465.57	398.37	984.85	638.86	1356.38	698.47	1057.94	1137.98
	Hungary	580.86	618.17	633.09	1548.58	1301.36	1790.80	1030.79	1260.97	1192.32
	Iceland	4470.35	1767.52	1683.62	4538.20	4267.90	4469.77	4027.78	3716.85	3916.83
	Ireland	927.30	1016.47	847.09	1228.21	1351.63	1497.71	888.25	1045.88	1140.08
	Israel	243.96	226.55	255.00				763.57	748.70	944.48
	Italy	1146.90	1250.06	1087.14	2081.84	1719.60	1429.99	2393.53	2271.05	2465.11
	Japan	299.87	275.78	319.65	946.66	685.05	763.87	921.31	764.88	922.59
	Korea	934.90	1046.87	1094.97	886.94	994.35	896.94	788.83	666.43	848.69
	Luxembourg	4603.35	2414.63	1983.25	6121.63	9060.50	5889.69	7380.27	8698.35	8991.96
	Mexico	782.61	713.63	696.16	1013.38	875.71	3649.58	1711.95	2414.58	1532.76
	Netherlands	739.02	636.25	600.72	1137.22	948.97	1266.98	1483.84	1392.84	1430.80
	New Zealand	1559.79	1127.75	1810.89	1991.14	2286.99	2260.48	1447.27	1804.97	1653.84
	Norway	1457.05	1357.33	1278.66	1722.78	1545.02	1485.70	1205.01	1358.56	1008.31
	Poland	580.77	379.75	512.57	1302.30	1462.04	1327.67	1381.16	1603.00	1599.93
	Portugal	552.71	549.83	512.58	683.00	673.47	828.55	982.27	1173.32	1055.78
	Slovak Republic				907.63	883.26	776.01	1338.09	1604.60	1464.78
	Slovenia							9243.87	9015.32	8372.96
	Spain	1142.73	866.10	1082.62	2463.13	1838.05	2031.26	2100.02	3158.32	2387.91
	Sweden	2105.73	1608.72	1558.22	1820.82	1453.55	2190.55	1349.69	1474.78	1729.52
	Switzerland	607.29	618.19	656.11	1022.60	1079.70	1019.58	1233.95	1376.12	1120.58
	Turkey				337.29	300.61	333.61	609.34	479.78	484.81
	United Kingdom	1681.71	1569.58	1687.50	2137.52	1816.71	1983.80	2476.48	2050.31	3079.09
	United States	243.09	181.38	214.94	1462.29	1418.13	1436.93		570.69	651.84
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	977.17	1410.39	1427.45						
	Argentina	143.68	193.87	214.32				387.98	349.52	308.84
	Azerbaijan							800.36	573.83	494.38
	Brazil	920.17	864.14	1253.39	810.46	521.17	956.42	1199.91	1192.87	1430.69
	Bulgaria	488.19	385.72	585.59				316.65	331.67	354.30
	Colombia							609.76	598.30	920.04
	Costa Rica									
	Croatia							1177.22	1389.31	1373.91
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	863.08	906.81	892.81	568.38	691.11	578.42	1237.17	1383.92	1422.27
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	488.54	431.65	468.13	1006.67	619.27	762.17	206.00	391.62	173.31
	Jordan							1248.74	768.90	1076.03
	Kazakhstan									
	Kyrgyzstan							1011.87	754.13	845.56
	Latvia	451.43	631.95	317.17	729.53	670.91	653.88	674.99	787.34	870.45
	Liechtenstein	600.15	216.44	184.60	663.74	699.92	666.34	648.78	592.87	629.78
	Lithuania							1144.10	1217.22	1115.47
	Macao-China				1238.82	956.00	1002.09	5857.30	5820.12	5947.11
	Macedonia	2909.15	1587.55	1650.28						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							5938.26	4837.29	6226.46
	Panama									
	Peru	522.64	710.82	1022.40						
	Qatar							10253.67	10257.33	10790.51
	Romania	1086.23	838.80	904.25				534.63	553.27	397.54
	Russian Federation	568.47	417.94	501.21	686.70	618.36	669.62	658.88	659.51	646.74
	Russian Federation (Perm)									
	Serbia				580.01	654.29	759.02	800.13	905.89	823.72
	Shanghai-China									
	Singapore									
	Chinese Taipei							995.46	747.73	746.74
	Thailand	632.89	647.84	693.95	1319.89	936.69	1205.39	1189.06	1537.21	1403.42
	Trinidad and Tobago									
	Tunisia				1725.36	1097.17	1282.09	643.26	643.21	795.36
	United Arab Emirates									
	Uruguay				1683.08	1012.20	1478.10	1443.82	1734.32	1329.05
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 11.13 Effective sample size 2 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1926.35	1747.90	1716.27	4279.76	3936.42	4022.31
	Austria	1315.46	1468.17	1064.03	1257.65	1271.35	1437.77
	Belgium	2164.85	2610.74	2179.49	2393.00	2731.09	2486.37
	Canada	4135.66	4630.65	3835.75	3196.04	2762.78	2381.39
	Chile	902.28	782.22	886.70	861.32	874.65	868.90
	Czech Republic	1253.49	1203.89	1173.14	1334.12	1492.31	1522.60
	Denmark	1809.70	1219.99	1805.59	1137.23	1345.28	1181.27
	Estonia	1210.78	1162.47	1265.63	1689.95	1905.77	1882.94
	Finland	1544.41	1464.81	1660.54	1933.31	2249.62	2100.84
	France	975.07	1164.32	855.69	1617.90	2020.71	1603.04
	Germany	1371.02	1418.78	1374.39	1113.72	1166.14	1072.10
	Greece	636.31	642.08	822.82	1004.79	1410.04	910.87
	Hungary	810.01	761.89	796.01	985.42	1038.99	1061.32
	Iceland	4792.00	4846.09	4689.90	3352.93	3269.26	3346.34
	Ireland	1045.57	1161.71	1019.50	1268.64	1558.74	1677.71
	Israel	1185.47	1071.69	1512.55	634.03	543.59	566.49
	Italy	4584.41	3037.69	4502.38	2902.38	2633.83	3532.05
	Japan	889.00	868.03	948.08	766.64	697.76	863.69
	Korea	663.57	521.11	826.92	547.13	522.20	551.69
	Luxembourg	8367.52	8390.23	8694.32	7779.48	9796.40	8692.14
	Mexico	2609.34	1917.44	3142.19	3353.75	3344.23	3236.71
	Netherlands	338.85	377.75	380.69	733.36	878.29	773.42
	New Zealand	2074.21	1911.97	1825.02	2124.41	2324.64	2446.67
	Norway	1503.43	1426.82	1436.94	1283.75	1138.82	1081.75
	Poland	1311.37	1172.09	1679.64	860.86	638.28	932.98
	Portugal	850.92	1137.41	968.11	683.91	723.78	677.21
	Slovak Republic	1377.80	1002.53	1126.79	759.59	973.45	902.43
	Slovenia	8350.69	7214.62	7798.98	7641.21	6916.45	6752.19
	Spain	2061.81	3110.77	2334.60	2903.97	3063.03	2611.89
	Sweden	1262.14	1214.77	1418.99	1307.75	1744.55	1200.33
	Switzerland	1682.23	1636.39	1481.42	1427.12	1177.77	1526.94
	Turkey	627.22	472.10	578.47	452.85	416.69	503.66
	United Kingdom	1905.53	1550.95	1830.77	938.47	1042.87	1049.75
	United States	767.89	689.42	799.56	668.46	866.49	676.10
	United States (Connecticut)[1]				293.84	270.05	387.18
	United States (Florida)[1]				214.72	232.62	216.73
	United States (Massachusetts)[1]				295.98	257.35	282.45
Partners	Albania	631.63	716.96	578.40	1643.05	2641.23	1725.62
	Argentina	596.48	553.50	511.96	740.52	499.51	558.01
	Azerbaijan	692.85	612.11	815.51			
	Brazil	1313.65	1519.73	1603.58	1896.97	1753.45	2162.52
	Bulgaria	344.37	296.16	342.62	423.11	586.97	483.60
	Colombia	593.61	804.44	628.47	706.63	697.11	791.81
	Costa Rica	720.59	685.15	915.78	743.68	605.35	889.25
	Croatia	966.99	961.79	1019.12	775.36	635.68	804.89
	Cyprus[2, 3]				9143.79	7801.96	7354.41
	Georgia	1418.56	1035.22	1075.37			
	Hong Kong-China	1702.91	1284.18	1021.26	1089.85	1024.67	1119.54
	India (Himachal Pradesh)	442.95	339.58	311.57			
	India (Tamil Nadu)	222.44	242.05	265.71			
	Indonesia	393.38	431.73	482.48	415.74	428.43	680.20
	Jordan	780.52	541.11	726.32	746.29	648.36	779.19
	Kazakhstan	1010.49	946.10	833.06	1041.35	686.26	900.22
	Kyrgyzstan	997.17	854.77	1144.58			
	Latvia	873.43	718.79	910.05	1358.45	982.69	978.37
	Liechtenstein	919.62	531.92	697.09	542.06	634.06	630.72
	Lithuania	1425.96	1243.37	945.73	1258.17	1259.95	1309.08
	Macao-China	9031.40	9394.29	7905.60	9016.56	9730.28	9057.67
	Macedonia						
	Malaysia	834.12	862.99	1008.91	753.51	818.33	889.05
	Malta	5925.31	5572.04	5409.61			
	Mauritius	9004.72	9371.20	8001.39			
	Moldova	1241.42	894.35	970.20			
	Montenegro	3212.90	2112.12	2421.72	7152.93	7158.19	7033.25
	Panama	332.84	277.92	264.29			
	Peru	704.36	1019.78	1213.94	572.02	570.06	656.67
	Qatar	22891.89	20465.48	16870.28	20512.73	19180.52	20695.63
	Romania	489.11	708.65	599.50	617.44	667.10	635.70
	Russian Federation	829.39	706.68	909.06	1124.73	881.92	987.36
	Russian Federation (Perm)				276.06	303.41	319.98
	Serbia	1344.06	989.32	1312.34	859.49	772.61	782.87
	Shanghai-China	1264.94	1408.35	1519.34	1211.47	1136.25	1060.81
	Singapore	9141.26	7408.78	7198.54	5735.82	6668.53	5287.92
	Chinese Taipei	1249.38	998.37	1269.23	1075.24	1375.23	1397.47
	Thailand	846.91	612.75	911.86	775.87	630.02	836.24
	Trinidad and Tobago	8514.55	8120.71	8694.13			
	Tunisia	948.07	744.55	1010.97	457.09	432.34	625.66
	United Arab Emirates	1475.68	1726.61	1944.06	1679.14	1668.91	1194.82
	Uruguay	1643.20	1718.68	2013.78	1023.22	1090.89	1237.88
	Venezuela (Miranda)	383.31	414.30	463.33			
	Viet Nam				297.81	325.24	348.43

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.14 Design effect 3 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	5.90	3.81	3.67	5.77	6.25	5.19	6.69	9.08	7.09
	Austria	3.10	1.93	2.01	6.02	5.52	5.69	6.91	6.96	7.81
	Belgium	7.31	4.98	5.53	4.73	3.75	4.20	6.70	6.84	5.44
	Canada	7.97	4.42	5.06	10.39	11.67	10.75	14.24	11.82	10.40
	Chile	7.66	3.86	3.20				12.08	14.09	11.53
	Czech Republic	3.18	2.51	1.97	7.96	8.42	6.54	8.27	6.88	7.34
	Denmark	2.36	1.65	1.70	3.90	3.57	3.30	5.58	4.12	4.88
	Estonia							5.69	5.60	4.28
	Finland	3.90	1.68	1.99	2.22	2.63	2.33	3.17	3.19	2.39
	France	4.02	2.19	2.26	3.12	3.09	2.87	7.15	4.99	5.14
	Germany	2.36	1.65	1.41	4.44	4.86	4.84	7.52	7.31	6.96
	Greece	12.04	6.68	6.60	6.60	7.89	5.72	7.48	4.94	5.59
	Hungary	8.64	4.66	4.58	3.32	4.19	3.41	5.18	4.24	4.04
	Iceland	0.73	1.08	1.11	0.70	0.77	0.74	0.90	1.03	0.96
	Ireland	4.50	2.17	2.55	3.44	3.08	2.99	6.41	5.08	4.92
	Israel	26.61	12.44	12.82				6.63	7.28	5.02
	Italy	4.90	2.59	2.62	9.72	11.24	9.59	10.64	12.07	9.62
	Japan	19.28	11.57	10.50	6.20	7.42	6.66	7.39	7.99	6.71
	Korea	5.89	2.84	2.85	7.39	6.47	6.63	9.18	8.44	7.01
	Luxembourg	0.73	0.79	0.98	0.51	0.43	0.58	0.54	0.46	0.48
	Mexico	6.69	4.06	4.15	54.56	53.89	43.63	30.91	34.61	34.30
	Netherlands	3.52	2.27	2.35	4.23	4.48	3.78	4.10	3.96	3.83
	New Zealand	2.40	1.93	1.12	2.39	2.17	2.15	3.73	2.98	3.00
	Norway	2.97	1.87	1.85	2.72	2.68	2.98	4.19	3.77	4.86
	Poland	7.12	5.56	5.28	3.77	3.25	3.39	4.24	4.14	3.68
	Portugal	9.72	4.98	5.11	7.36	6.94	6.19	6.36	5.51	5.72
	Slovak Republic				8.33	9.31	9.66	3.87	3.79	3.52
	Slovenia							0.67	0.67	0.77
	Spain	6.18	4.04	3.27	7.19	7.64	6.96	12.06	12.34	14.82
	Sweden	2.32	1.59	1.64	2.80	3.31	2.59	4.79	3.14	2.72
	Switzerland	10.52	6.37	6.40	9.85	9.68	9.69	12.60	12.33	12.22
	Turkey				17.67	19.84	18.03	9.88	13.33	10.49
	United Kingdom	5.97	3.70	3.61	6.08	6.34	5.56	6.23	7.45	5.63
	United States	17.29	12.79	11.01	5.05	4.87	4.69		11.11	8.87
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	5.38	2.14	2.27						
	Argentina	32.64	13.32	13.21				14.17	16.20	15.54
	Azerbaijan							9.66	11.22	12.47
	Brazil	6.14	3.68	2.90	7.17	10.23	7.83	11.80	9.23	8.66
	Bulgaria	10.63	6.97	5.49				15.44	16.32	14.58
	Colombia							9.60	8.16	6.63
	Costa Rica									
	Croatia							5.03	4.08	4.12
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	5.31	2.85	2.93	8.39	8.76	8.99	3.99	3.66	3.35
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	21.83	11.49	10.96	20.17	24.89	23.28	66.45	51.69	71.00
	Jordan							7.35	9.94	6.42
	Kazakhstan									
	Kyrgyzstan							6.67	8.91	7.19
	Latvia	10.16	3.83	7.08	7.42	7.96	7.98	7.84	6.26	5.78
	Liechtenstein	0.48	0.78	0.95	0.47	0.36	0.42	0.48	0.48	0.48
	Lithuania							4.51	4.74	4.40
	Macao-China				1.01	1.32	1.29	0.77	0.75	0.78
	Macedonia	1.68	1.71	1.56						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							0.73	0.90	0.69
	Panama									
	Peru	9.24	3.93	3.84						
	Qatar							0.52	0.49	0.53
	Romania	5.31	3.51	3.27				12.96	12.47	13.65
	Russian Federation	13.53	10.09	8.34	10.41	12.09	10.14	12.06	10.59	9.63
	Russian Federation (Perm)									
	Serbia				8.30	8.38	7.52	6.69	6.70	6.06
	Shanghai-China									
	Singapore									
	Chinese Taipei							13.51	13.77	12.52
	Thailand	9.40	5.39	4.60	6.06	6.75	5.45	6.02	4.83	4.69
	Trinidad and Tobago									
	Tunisia				3.58	4.47	3.96	7.82	8.41	5.96
	United Arab Emirates									
	Uruguay				4.31	6.24	4.07	3.73	3.14	3.98
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.14 Design effect 3 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	7.88	10.05	8.80	3.69	4.07	4.21
	Austria	5.57	4.92	6.61	4.16	3.90	3.85
	Belgium	4.38	3.76	4.65	4.02	3.58	3.84
	Canada	6.14	7.31	7.31	9.01	9.06	9.62
	Chile	7.81	8.12	7.16	9.31	9.63	8.61
	Czech Republic	5.72	5.48	5.56	5.18	4.46	5.07
	Denmark	3.53	5.21	4.00	7.07	5.79	6.52
	Estonia	4.52	4.57	4.48	2.97	2.79	2.71
	Finland	3.90	4.00	3.84	5.37	4.41	4.76
	France	4.53	3.95	5.24	3.01	2.64	3.00
	Germany	3.83	4.02	3.79	4.69	4.42	4.81
	Greece	9.94	9.14	9.07	5.51	4.02	6.25
	Hungary	5.70	6.40	6.02	5.52	5.38	5.01
	Iceland	0.75	0.72	0.77	1.05	1.08	1.07
	Ireland	3.82	3.45	4.30	4.28	3.44	3.40
	Israel	5.86	5.65	4.62	9.47	9.98	10.50
	Italy	8.07	12.09	9.89	12.61	14.55	12.93
	Japan	7.20	7.53	7.02	8.73	9.32	8.81
	Korea	9.26	10.05	8.28	10.30	10.63	9.93
	Luxembourg	0.45	0.45	0.43	0.46	0.40	0.45
	Mexico	19.99	20.44	19.81	11.82	11.11	11.43
	Netherlands	15.91	13.44	14.91	6.18	6.14	6.01
	New Zealand	2.34	2.58	2.63	2.12	1.96	1.77
	Norway	3.53	3.55	3.72	4.48	4.24	4.45
	Poland	4.08	4.86	3.50	5.83	7.36	5.80
	Portugal	7.80	6.25	7.43	9.10	9.23	9.99
	Slovak Republic	3.53	4.65	4.36	7.27	5.27	5.80
	Slovenia	0.72	0.82	0.76	0.70	0.82	0.83
	Spain	13.70	13.33	13.97	10.63	11.39	11.19
	Sweden	3.82	4.20	3.34	3.71	2.81	4.21
	Switzerland	7.92	12.24	9.90	8.94	11.45	9.64
	Turkey	9.06	11.23	9.77	11.57	13.46	11.33
	United Kingdom	6.87	9.19	7.72	16.19	15.13	14.34
	United States	7.39	8.03	7.18	8.13	7.59	7.95
	United States (Connecticut)[1]				7.01	6.57	5.45
	United States (Florida)[1]				9.05	8.74	9.08
	United States (Massachusetts)[1]				6.63	6.92	6.51
Partners	Albania	7.48	8.45	8.90	3.36	2.01	2.85
	Argentina	8.65	9.08	9.57	8.67	12.44	11.90
	Azerbaijan	8.77	8.61	7.60			
	Brazil	16.84	17.10	16.53	10.55	11.70	12.47
	Bulgaria	15.49	15.75	13.83	13.54	9.47	11.50
	Colombia	14.69	14.14	15.57	15.20	13.63	14.24
	Costa Rica	7.11	7.73	6.03	9.55	8.91	7.44
	Croatia	5.33	5.95	5.44	7.24	8.00	6.53
	Cyprus[2, 3]				0.54	0.64	0.66
	Georgia	3.81	4.84	4.71			
	Hong Kong-China	2.99	3.91	4.78	4.83	5.07	4.47
	India (Himachal Pradesh)	4.18	5.39	5.22			
	India (Tamil Nadu)	15.66	17.27	13.75			
	Indonesia	15.97	14.17	15.08	17.20	17.61	16.43
	Jordan	8.57	12.97	10.08	10.52	11.31	9.91
	Kazakhstan	6.01	6.92	6.99	7.32	10.25	8.90
	Kyrgyzstan	5.16	6.22	4.99			
	Latvia	5.96	6.68	6.54	3.33	4.76	5.07
	Liechtenstein	0.33	0.56	0.43	0.46	0.42	0.42
	Lithuania	3.37	3.90	5.24	3.77	3.95	3.94
	Macao-China	0.58	0.60	0.64	0.55	0.54	0.57
	Macedonia						
	Malaysia	6.24	6.67	5.99	8.04	7.73	7.27
	Malta	0.56	0.60	0.57			
	Mauritius	0.46	0.44	0.50			
	Moldova	4.98	6.77	6.13			
	Montenegro	1.55	2.55	2.30	0.60	0.61	0.64
	Panama	16.78	16.64	16.04			
	Peru	9.53	10.81	8.30	12.78	11.45	12.25
	Qatar	0.39	0.41	0.42	0.49	0.53	0.52
	Romania	9.87	8.53	8.58	9.65	10.42	8.57
	Russian Federation	7.19	7.92	6.90	5.41	6.37	5.79
	Russian Federation (Perm)				7.05	6.54	6.61
	Serbia	4.53	5.71	4.33	6.28	6.46	6.94
	Shanghai-China	4.44	3.78	3.86	6.05	5.30	6.64
	Singapore	0.55	0.60	0.67	0.97	0.82	1.05
	Chinese Taipei	5.15	6.06	5.22	6.47	4.80	4.66
	Thailand	8.26	10.35	8.43	10.08	11.51	9.49
	Trinidad and Tobago	0.55	0.45	0.49			
	Tunisia	5.58	7.24	5.30	11.37	10.96	8.29
	United Arab Emirates	8.81	7.53	7.79	7.74	8.22	10.24
	Uruguay	3.97	4.40	3.80	5.66	5.10	4.43
	Venezuela (Miranda)	7.91	7.40	7.19			
	Viet Nam				17.41	15.75	15.32

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.15 Effective sample size 3 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	877.35	751.27	778.69	2175.77	2006.59	2417.33	2117.65	1560.44	1998.74
	Austria	1530.53	1365.29	1326.79	764.10	832.95	808.29	713.08	707.88	631.23
	Belgium	912.36	760.51	673.60	1860.76	2348.66	2093.47	1322.55	1295.13	1627.30
	Canada	3726.10	3726.41	3259.95	2689.72	2395.52	2600.93	1590.68	1916.06	2176.50
	Chile	637.84	705.84	847.37				433.37	371.51	453.94
	Czech Republic	1687.66	1220.71	1554.17	794.12	750.75	966.00	717.06	861.71	807.76
	Denmark	1795.91	1440.39	1383.00	1080.85	1182.29	1279.16	812.28	1098.70	928.93
	Estonia							854.85	868.81	1136.58
	Finland	1245.76	1609.89	1362.71	2609.23	2204.28	2491.84	1486.36	1476.78	1973.32
	France	1163.85	1183.99	1148.50	1379.61	1392.56	1498.13	659.29	945.69	917.91
	Germany	2151.65	1710.95	2031.29	1049.62	959.12	962.78	650.52	668.75	702.29
	Greece	388.03	390.07	393.15	700.69	586.41	809.56	651.79	986.24	871.80
	Hungary	565.73	601.21	611.60	1436.54	1137.70	1395.49	866.41	1057.75	1112.25
	Iceland	4632.69	1741.12	1678.79	4774.12	4337.83	4551.74	4190.88	3676.91	3932.57
	Ireland	856.14	979.45	838.29	1128.02	1258.18	1296.33	715.46	902.79	931.07
	Israel	169.06	199.57	196.09				691.79	629.60	912.46
	Italy	1017.92	1065.73	1053.85	1196.85	1035.38	1213.06	2046.48	1804.14	2262.56
	Japan	272.63	252.64	277.48	759.10	634.79	707.20	805.44	744.81	886.66
	Korea	846.25	973.54	967.81	736.93	841.94	820.53	563.84	613.11	737.90
	Luxembourg	4838.18	2479.56	1988.20	7655.42	9220.10	6738.79	8461.35	9884.01	9610.18
	Mexico	687.91	632.82	613.41	549.50	556.41	687.15	1002.04	894.73	902.85
	Netherlands	710.58	609.97	592.87	943.84	890.65	1057.29	1187.17	1229.33	1272.70
	New Zealand	1530.53	1060.11	1805.46	1886.10	2076.56	2094.08	1293.39	1618.51	1609.43
	Norway	1398.41	1234.43	1246.35	1495.23	1516.95	1365.62	1119.13	1243.78	965.13
	Poland	512.97	355.61	386.72	1163.73	1349.14	1293.28	1308.79	1339.30	1507.10
	Portugal	471.76	511.44	499.46	626.44	664.21	744.70	803.42	927.82	892.79
	Slovak Republic				882.13	788.89	760.54	1223.20	1248.78	1345.71
	Slovenia							9871.79	9837.00	8541.07
	Spain	1005.47	847.53	1057.03	1501.87	1412.79	1549.89	1624.90	1589.12	1322.63
	Sweden	1902.83	1546.07	1488.16	1652.85	1396.00	1788.24	928.50	1415.26	1631.23
	Switzerland	580.11	533.11	530.51	855.12	869.97	868.93	967.72	988.69	997.36
	Turkey				274.82	244.73	269.30	500.41	370.82	471.20
	United Kingdom	1563.58	1405.56	1432.97	1567.01	1504.34	1715.93	2111.82	1766.26	2336.77
	United States	222.46	166.87	193.39	1081.03	1120.42	1164.18		504.84	632.78
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	925.00	1301.48	1223.80						
	Argentina	122.02	167.46	167.32				306.24	267.84	279.30
	Azerbaijan							536.53	461.97	415.83
	Brazil	797.21	738.81	935.14	621.23	435.30	568.72	787.54	1006.74	1073.56
	Bulgaria	437.99	375.35	463.56				291.37	275.62	308.45
	Colombia							466.60	548.57	675.02
	Costa Rica									
	Croatia							1037.09	1277.81	1264.80
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	830.27	854.87	830.98	534.02	511.25	497.91	1164.50	1267.56	1388.58
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	337.45	355.97	371.85	533.39	432.28	462.33	160.22	205.99	149.95
	Jordan							885.81	654.70	1014.26
	Kazakhstan									
	Kyrgyzstan							884.73	662.37	821.11
	Latvia	383.16	561.75	304.74	623.64	581.24	579.66	602.25	754.02	816.85
	Liechtenstein	658.15	224.26	184.94	699.33	911.11	797.95	712.53	710.00	709.40
	Lithuania							1052.35	1000.95	1077.47
	Macao-China				1236.24	944.67	966.91	6150.53	6374.04	6079.41
	Macedonia	2678.63	1485.03	1617.04						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							6113.64	4942.66	6492.50
	Panama									
	Peru	479.58	626.35	640.48						
	Qatar							12151.18	12696.73	11899.81
	Romania	910.11	764.95	823.87				394.82	410.39	375.01
	Russian Federation	495.39	368.60	445.85	573.92	494.29	589.17	480.89	547.42	602.30
	Russian Federation (Perm)									
	Serbia				530.42	525.94	585.96	717.53	716.27	792.19
	Shanghai-China									
	Singapore									
	Chinese Taipei							652.52	640.18	704.30
	Thailand	567.91	549.28	645.13	864.59	775.34	960.64	1029.30	1282.48	1319.38
	Trinidad and Tobago									
	Tunisia				1319.83	1056.67	1193.43	593.10	551.59	779.09
	United Arab Emirates									
	Uruguay				1352.83	935.41	1434.75	1298.87	1541.35	1216.86
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.15 Effective sample size 3 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1807.95	1417.93	1620.17	3925.12	3558.30	3441.68
	Austria	1183.01	1339.41	997.64	1141.97	1219.69	1236.61
	Belgium	1942.50	2261.20	1829.81	2136.03	2401.27	2238.90
	Canada	3779.99	3173.53	3175.41	2389.91	2376.71	2238.61
	Chile	725.56	697.84	792.07	736.32	712.09	796.15
	Czech Republic	1059.55	1106.46	1089.91	1028.14	1194.14	1050.71
	Denmark	1676.86	1137.80	1479.77	1058.61	1291.16	1148.05
	Estonia	1045.23	1035.00	1055.83	1608.34	1710.78	1762.92
	Finland	1489.32	1453.17	1511.86	1645.46	2004.10	1853.40
	France	948.03	1088.63	820.31	1533.13	1744.36	1539.86
	Germany	1298.79	1239.46	1313.80	1066.78	1131.17	1040.07
	Greece	500.04	543.76	547.97	930.55	1275.84	819.89
	Hungary	807.69	719.80	764.48	871.66	893.33	959.49
	Iceland	4842.80	5062.70	4724.40	3333.51	3242.65	3269.37
	Ireland	1030.53	1141.84	914.85	1170.80	1457.76	1474.31
	Israel	983.56	1019.80	1248.15	533.93	506.37	481.27
	Italy	3827.52	2557.02	3123.54	2464.18	2135.70	2403.29
	Japan	845.71	808.58	867.39	727.12	681.75	720.78
	Korea	538.85	496.49	602.50	488.46	473.30	506.59
	Luxembourg	10200.87	10290.01	10632.06	11508.90	13053.99	11662.98
	Mexico	1913.21	1871.45	1930.62	2859.78	3044.01	2957.87
	Netherlands	299.10	354.20	319.24	721.62	725.88	742.40
	New Zealand	1980.05	1802.26	1767.80	2025.88	2185.40	2423.32
	Norway	1320.17	1312.83	1254.36	1045.65	1106.11	1051.91
	Poland	1206.17	1011.20	1405.87	789.61	626.17	793.66
	Portugal	807.69	1008.13	847.93	628.45	619.95	572.90
	Slovak Republic	1291.73	978.55	1045.86	643.74	886.85	806.72
	Slovenia	8585.28	7461.18	8149.83	8484.51	7225.16	7113.94
	Spain	1890.14	1942.00	1853.22	2380.37	2222.54	2261.67
	Sweden	1196.83	1087.43	1368.62	1275.90	1687.00	1125.63
	Switzerland	1490.68	965.15	1192.68	1256.20	980.67	1164.78
	Turkey	551.13	444.86	511.39	418.90	360.20	428.01
	United Kingdom	1771.89	1324.85	1577.11	781.66	836.80	882.96
	United States	708.55	651.30	728.94	612.21	656.15	625.83
	United States (Connecticut)[1]				242.10	258.15	311.31
	United States (Florida)[1]				209.44	216.96	208.80
	United States (Massachusetts)[1]				259.90	248.90	264.73
Partners	Albania	614.63	543.66	516.65	1411.99	2364.46	1664.35
	Argentina	551.63	525.75	498.69	681.18	475.08	496.63
	Azerbaijan	534.81	545.03	617.38			
	Brazil	1195.15	1177.15	1217.58	1819.56	1641.59	1540.34
	Bulgaria	291.01	286.14	325.83	390.01	557.90	459.22
	Colombia	539.25	560.36	508.80	596.82	665.87	637.31
	Costa Rica	644.27	592.58	758.82	481.83	516.55	618.46
	Croatia	937.55	839.08	918.05	692.08	626.39	766.77
	Cyprus[2, 3]				9414.04	7966.99	7708.08
	Georgia	1220.22	959.61	986.01			
	Hong Kong-China	1617.79	1236.90	1011.49	966.66	921.92	1044.75
	India (Himachal Pradesh)	386.39	299.85	309.50			
	India (Tamil Nadu)	204.98	185.84	233.48			
	Indonesia	321.54	362.44	340.63	326.85	319.23	342.19
	Jordan	756.61	499.92	643.55	668.80	622.37	710.51
	Kazakhstan	899.94	781.86	774.71	792.94	566.49	652.29
	Kyrgyzstan	965.69	801.13	999.49			
	Latvia	755.22	674.02	687.91	1293.98	905.24	849.97
	Liechtenstein	998.86	582.31	757.61	635.53	705.47	695.85
	Lithuania	1342.49	1161.50	863.47	1226.44	1169.45	1173.46
	Macao-China	10304.69	9856.54	9283.65	9635.73	9843.77	9391.35
	Macedonia						
	Malaysia	801.46	749.56	835.18	646.23	672.66	714.62
	Malta	6128.34	5762.83	6072.09			
	Mauritius	10025.84	10629.68	9221.89			
	Moldova	1043.12	767.55	847.43			
	Montenegro	3111.98	1889.60	2098.26	7866.88	7808.95	7454.56
	Panama	236.58	238.52	247.38			
	Peru	628.30	553.87	721.28	472.30	526.87	492.59
	Qatar	23068.35	21955.25	21388.79	22328.78	20708.44	20987.67
	Romania	484.11	560.13	556.47	525.85	487.13	592.31
	Russian Federation	738.03	670.51	769.64	966.70	821.17	902.70
	Russian Federation (Perm)				249.72	269.28	266.23
	Serbia	1220.42	967.20	1274.69	746.22	725.06	674.74
	Shanghai-China	1151.87	1353.71	1325.33	855.23	976.71	780.06
	Singapore	9655.76	8867.12	7869.82	5747.04	6737.82	5262.07
	Chinese Taipei	1133.02	962.66	1117.95	934.97	1259.80	1297.23
	Thailand	753.70	601.48	738.31	655.10	573.87	696.33
	Trinidad and Tobago	8643.87	10722.07	9800.46			
	Tunisia	888.06	684.85	935.42	387.55	401.99	531.74
	United Arab Emirates	1233.69	1443.92	1395.15	1485.71	1399.27	1123.25
	Uruguay	1501.61	1355.13	1566.22	938.58	1042.07	1198.83
	Venezuela (Miranda)	366.83	391.93	403.35			
	Viet Nam				284.79	314.78	323.60

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]
Table 11.16 Design effect 4 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1.05	1.13	1.06	1.04	1.02	1.03	1.02	1.01	1.02
	Austria	1.02	1.00	1.03	1.02	1.03	1.02	1.01	1.03	1.02
	Belgium	1.01	1.03	1.01	1.03	1.02	1.11	1.01	1.00	1.01
	Canada	1.01	1.03	1.02	1.05	1.04	1.08	1.02	1.01	1.01
	Chile	1.02	1.09	1.12				1.01	1.02	1.01
	Czech Republic	1.02	1.01	1.04	1.04	1.03	1.09	1.01	1.02	1.01
	Denmark	1.03	1.14	1.02	1.10	1.07	1.09	1.03	1.05	1.03
	Estonia							1.01	1.01	1.03
	Finland	1.04	1.15	1.12	1.07	1.10	1.12	1.04	1.19	1.10
	France	1.03	1.09	1.11	1.05	1.04	1.09	1.01	1.04	1.00
	Germany	1.06	1.03	1.16	1.01	1.00	1.03	1.01	1.02	1.01
	Greece	1.02	1.04	1.00	1.08	1.01	1.17	1.01	1.02	1.07
	Hungary	1.00	1.01	1.01	1.03	1.05	1.13	1.05	1.06	1.02
	Iceland	1.15	1.23	1.03	1.20	1.08	1.07	1.69	1.55	1.12
	Ireland	1.02	1.03	1.01	1.04	1.04	1.08	1.05	1.04	1.06
	Israel	1.02	1.01	1.03				1.02	1.03	1.01
	Italy	1.03	1.12	1.02	1.09	1.07	1.02	1.02	1.02	1.01
	Japan	1.01	1.01	1.02	1.05	1.01	1.01	1.02	1.00	1.01
	Korea	1.02	1.04	1.08	1.03	1.03	1.02	1.05	1.01	1.03
	Luxembourg	1.22	1.14	1.15	1.71	1.03	1.44	1.39	1.29	1.14
	Mexico	1.02	1.04	1.04	1.02	1.01	1.11	1.02	1.05	1.02
	Netherlands	1.02	1.04	1.01	1.07	1.02	1.08	1.09	1.05	1.05
	New Zealand	1.01	1.07	1.02	1.04	1.09	1.07	1.05	1.06	1.01
	Norway	1.02	1.13	1.03	1.10	1.01	1.05	1.02	1.03	1.01
	Poland	1.02	1.02	1.08	1.05	1.04	1.01	1.02	1.07	1.02
	Portugal	1.02	1.02	1.01	1.01	1.00	1.02	1.04	1.06	1.04
	Slovak Republic				1.00	1.01	1.00	1.03	1.11	1.04
	Slovenia							1.24	1.34	1.10
	Spain	1.03	1.01	1.01	1.12	1.05	1.06	1.03	1.10	1.06
	Sweden	1.09	1.07	1.08	1.06	1.02	1.17	1.14	1.02	1.04
	Switzerland	1.00	1.03	1.05	1.02	1.03	1.02	1.02	1.04	1.01
	Turkey				1.01	1.01	1.01	1.03	1.02	1.00
	United Kingdom	1.02	1.05	1.07	1.08	1.04	1.04	1.03	1.03	1.07
	United States	1.01	1.01	1.01	1.10	1.07	1.07		1.01	1.00
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	1.01	1.08	1.15						
	Argentina	1.01	1.01	1.02				1.02	1.02	1.01
	Azerbaijan							1.06	1.02	1.02
	Brazil	1.03	1.07	1.22	1.05	1.02	1.11	1.05	1.02	1.05
	Bulgaria	1.01	1.00	1.06				1.01	1.01	1.01
	Colombia							1.04	1.01	1.07
	Costa Rica									
	Croatia							1.03	1.03	1.03
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	1.01	1.03	1.04	1.01	1.05	1.02	1.02	1.04	1.01
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	1.02	1.02	1.03	1.05	1.02	1.03	1.00	1.02	1.00
	Jordan							1.07	1.02	1.01
	Kazakhstan									
	Kyrgyzstan							1.03	1.02	1.00
	Latvia	1.02	1.05	1.01	1.03	1.02	1.02	1.02	1.01	1.01
	Liechtenstein	1.20	1.19	1.04	1.11	1.58	1.40	1.21	1.47	1.28
	Lithuania							1.03	1.06	1.01
	Macao-China				1.29	1.04	1.15	1.27	1.53	1.11
	Macedonia	1.15	1.11	1.04						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							1.12	1.28	1.15
	Panama									
	Peru	1.01	1.05	1.27						
	Qatar							1.48	1.62	1.25
	Romania	1.05	1.04	1.05				1.03	1.03	1.00
	Russian Federation	1.01	1.01	1.02	1.02	1.02	1.02	1.03	1.02	1.01
	Russian Federation (Perm)									
	Serbia				1.01	1.03	1.05	1.02	1.05	1.01
	Shanghai-China									
	Singapore									
	Chinese Taipei							1.04	1.01	1.01
	Thailand	1.01	1.04	1.02	1.12	1.04	1.06	1.03	1.05	1.02
	Trinidad and Tobago									
	Tunisia				1.14	1.01	1.03	1.01	1.02	1.00
	United Arab Emirates									
	Uruguay				1.08	1.02	1.01	1.04	1.06	1.03
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 11.16 Design effect 4 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1.01	1.03	1.01	1.03	1.04	1.06
	Austria	1.03	1.03	1.01	1.03	1.01	1.06
	Belgium	1.04	1.06	1.06	1.04	1.06	1.04
	Canada	1.02	1.08	1.03	1.04	1.02	1.01
	Chile	1.04	1.02	1.02	1.02	1.03	1.01
	Czech Republic	1.04	1.02	1.02	1.08	1.08	1.12
	Denmark	1.03	1.02	1.08	1.01	1.01	1.01
	Estonia	1.05	1.04	1.06	1.03	1.07	1.04
	Finland	1.01	1.00	1.04	1.04	1.04	1.04
	France	1.01	1.02	1.01	1.03	1.11	1.02
	Germany	1.02	1.05	1.02	1.01	1.01	1.01
	Greece	1.03	1.02	1.07	1.02	1.04	1.02
	Hungary	1.00	1.01	1.01	1.03	1.04	1.03
	Iceland	1.04	1.18	1.03	1.13	1.11	1.48
	Ireland	1.01	1.01	1.04	1.03	1.03	1.06
	Israel	1.04	1.01	1.06	1.02	1.01	1.02
	Italy	1.03	1.02	1.05	1.02	1.02	1.04
	Japan	1.01	1.01	1.02	1.01	1.00	1.03
	Korea	1.03	1.01	1.05	1.01	1.01	1.01
	Luxembourg	1.49	1.51	1.48	2.49	1.73	1.87
	Mexico	1.02	1.00	1.03	1.02	1.01	1.01
	Netherlands	1.01	1.01	1.01	1.00	1.04	1.01
	New Zealand	1.04	1.04	1.02	1.05	1.07	1.01
	Norway	1.06	1.04	1.06	1.07	1.01	1.01
	Poland	1.03	1.04	1.08	1.02	1.00	1.04
	Portugal	1.01	1.03	1.02	1.01	1.02	1.02
	Slovak Republic	1.03	1.01	1.02	1.03	1.02	1.03
	Slovenia	1.11	1.23	1.22	1.49	1.31	1.43
	Spain	1.01	1.05	1.02	1.02	1.04	1.02
	Sweden	1.02	1.04	1.02	1.01	1.02	1.02
	Switzerland	1.02	1.07	1.03	1.02	1.02	1.04
	Turkey	1.02	1.01	1.02	1.01	1.01	1.02
	United Kingdom	1.01	1.02	1.02	1.01	1.02	1.01
	United States	1.01	1.01	1.02	1.01	1.05	1.01
	United States (Connecticut)[1]				1.04	1.01	1.06
	United States (Florida)[1]				1.00	1.01	1.00
	United States (Massachusetts)[1]				1.03	1.01	1.01
Partners	Albania	1.00	1.04	1.02	1.07	1.13	1.02
	Argentina	1.01	1.01	1.00	1.01	1.00	1.01
	Azerbaijan	1.04	1.02	1.05			
	Brazil	1.01	1.02	1.02	1.00	1.01	1.04
	Bulgaria	1.01	1.00	1.00	1.01	1.01	1.01
	Colombia	1.01	1.03	1.02	1.01	1.00	1.02
	Costa Rica	1.02	1.02	1.04	1.07	1.02	1.07
	Croatia	1.01	1.03	1.03	1.02	1.00	1.01
	Cyprus[2, 3]				1.07	1.06	1.16
	Georgia	1.06	1.02	1.03			
	Hong Kong-China	1.03	1.01	1.00	1.03	1.03	1.02
	India (Himachal Pradesh)	1.05	1.03	1.00			
	India (Tamil Nadu)	1.01	1.02	1.01			
	Indonesia	1.02	1.01	1.03	1.02	1.02	1.07
	Jordan	1.00	1.01	1.01	1.01	1.00	1.01
	Kazakhstan	1.03	1.04	1.01	1.05	1.02	1.05
	Kyrgyzstan	1.01	1.01	1.04			
	Latvia	1.03	1.01	1.06	1.02	1.02	1.04
	Liechtenstein	1.14	1.25	1.17	1.38	1.21	1.19
	Lithuania	1.03	1.02	1.02	1.01	1.03	1.04
	Macao-China	1.42	1.14	1.71	1.17	1.03	1.09
	Macedonia						
	Malaysia	1.01	1.03	1.04	1.02	1.03	1.04
	Malta	1.08	1.09	1.34			
	Mauritius	1.24	1.27	1.37			
	Moldova	1.05	1.03	1.03			
	Montenegro	1.07	1.09	1.14	1.30	1.27	1.18
	Panama	1.03	1.01	1.00			
	Peru	1.01	1.09	1.10	1.02	1.01	1.03
	Qatar	1.01	1.13	1.58	1.19	1.19	1.03
	Romania	1.00	1.04	1.01	1.02	1.04	1.01
	Russian Federation	1.02	1.01	1.03	1.04	1.01	1.02
	Russian Federation (Perm)				1.02	1.02	1.04
	Serbia	1.03	1.00	1.01	1.03	1.01	1.03
	Shanghai-China	1.03	1.01	1.05	1.09	1.04	1.07
	Singapore	1.14	1.69	1.35	1.06	1.06	1.10
	Chinese Taipei	1.03	1.01	1.03	1.03	1.02	1.02
	Thailand	1.02	1.00	1.03	1.02	1.01	1.02
	Trinidad and Tobago	1.03	1.78	1.28			
	Tunisia	1.02	1.01	1.02	1.02	1.01	1.02
	United Arab Emirates	1.03	1.03	1.06	1.02	1.03	1.01
	Uruguay	1.03	1.09	1.11	1.02	1.01	1.01
	Venezuela (Miranda)	1.01	1.01	1.02			
	Viet Nam				1.00	1.00	1.01

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.17 Effective sample size 4 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	4925.73	2533.61	2708.77	12098.41	12338.55	12230.66	13831.37	14010.18	13934.37
	Austria	4656.87	2629.63	2581.89	4524.91	4484.85	4523.93	4861.50	4796.10	4852.33
	Belgium	6617.13	3691.65	3701.79	8579.42	8655.09	7910.92	8761.71	8820.50	8762.04
	Canada	29363.81	16041.29	16181.44	26686.84	26790.25	25958.09	22182.87	22497.57	22367.34
	Chile	4814.61	2498.53	2420.44				5162.14	5130.69	5197.86
	Czech Republic	5250.88	3024.69	2945.61	6052.58	6165.61	5805.62	5858.86	5812.19	5885.03
	Denmark	4096.92	2089.94	2291.66	3833.17	3951.90	3872.23	4402.27	4332.87	4380.23
	Estonia							4802.09	4806.40	4705.25
	Finland	4696.56	2352.38	2414.20	5411.85	5287.31	5176.62	4539.54	3963.68	4300.83
	France	4542.28	2373.30	2336.75	4089.56	4143.39	3938.49	4680.02	4532.35	4696.38
	Germany	4799.60	2738.21	2465.69	4611.91	4648.17	4545.94	4845.17	4799.04	4833.23
	Greece	4599.79	2516.07	2586.85	4291.80	4566.87	3962.15	4819.07	4782.98	4548.57
	Hungary	4869.88	2777.35	2772.46	4604.14	4549.77	4204.80	4286.07	4224.06	4383.48
	Iceland	2936.25	1527.11	1809.02	2793.23	3112.52	3120.89	2246.07	2443.73	3372.33
	Ireland	3762.40	2059.14	2119.43	3738.53	3741.39	3577.31	4380.04	4406.43	4322.54
	Israel	4420.08	2453.64	2449.88				4499.42	4445.77	4543.98
	Italy	4821.84	2464.24	2712.13	10649.56	10887.49	11396.63	21389.53	21263.86	21546.71
	Japan	5227.27	2898.66	2867.36	4483.00	4648.87	4640.28	5817.78	5928.99	5909.73
	Korea	4875.01	2655.78	2560.67	5270.41	5263.67	5353.84	4923.28	5115.50	5046.69
	Luxembourg	2893.16	1712.80	1691.42	2300.53	3803.51	2730.43	3291.19	3542.30	3999.32
	Mexico	4488.51	2459.70	2438.58	29508.50	29656.33	26949.80	30235.91	29401.11	30321.61
	Netherlands	2463.24	1334.46	1382.33	3737.79	3916.85	3706.49	4478.14	4651.94	4656.74
	New Zealand	3616.71	1907.57	1979.87	4330.05	4120.07	4199.97	4612.63	4541.86	4756.30
	Norway	4058.37	2042.05	2236.77	3703.49	4019.21	3883.19	4579.09	4535.40	4637.53
	Poland	3574.93	1946.54	1887.71	4194.05	4219.83	4334.01	5452.15	5198.76	5419.44
	Portugal	4494.72	2496.92	2535.68	4542.40	4597.17	4507.82	4896.74	4808.78	4911.45
	Slovak Republic				7316.88	7239.95	7328.71	4575.67	4247.33	4564.52
	Slovenia							5322.26	4914.69	6021.91
	Spain	6050.17	3403.29	3420.06	9673.17	10301.39	10228.26	19084.62	17896.08	18461.04
	Sweden	4058.51	2295.42	2264.69	4362.36	4541.46	3966.48	3905.85	4355.15	4286.95
	Switzerland	6069.95	3295.01	3247.73	8230.22	8186.01	8250.77	11902.74	11769.84	12057.67
	Turkey				4788.67	4796.15	4786.85	4820.70	4824.15	4926.95
	United Kingdom	9197.93	4968.39	4826.00	8851.91	9163.82	9207.60	12717.20	12822.95	12247.95
	United States	3824.09	2119.18	2105.28	4980.32	5081.20	5109.09		5538.60	5589.51
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	4915.64	2577.45	2402.83						
	Argentina	3960.60	2200.74	2159.57				4250.97	4251.90	4307.42
	Azerbaijan							4889.51	5061.20	5098.56
	Brazil	4745.94	2543.82	2220.30	4231.87	4356.74	4004.63	8843.53	9086.18	8890.83
	Bulgaria	4601.46	2602.89	2396.82				4470.95	4438.28	4448.77
	Colombia							4317.63	4421.26	4188.62
	Costa Rica									
	Croatia							5037.91	5065.04	5068.68
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	4364.56	2357.86	2343.07	4438.96	4274.72	4387.37	4547.87	4484.74	4596.94
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	7209.55	4006.03	3970.01	10261.59	10565.63	10447.02	10600.45	10457.30	10623.22
	Jordan							6087.99	6381.93	6435.77
	Kazakhstan									
	Kyrgyzstan							5754.32	5800.56	5875.52
	Latvia	3817.27	2053.81	2142.45	4504.34	4524.35	4541.58	4635.33	4679.28	4653.89
	Liechtenstein	260.70	147.20	169.45	299.72	210.33	237.64	280.82	231.27	265.73
	Lithuania							4626.10	4469.05	4694.71
	Macao-China				968.77	1203.32	1089.19	3740.77	3104.25	4276.49
	Macedonia	3938.85	2297.50	2435.07						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							3983.01	3478.07	3871.70
	Panama									
	Peru	4380.65	2345.55	1943.79						
	Qatar							4236.25	3875.47	5024.52
	Romania	4611.27	2577.38	2577.18				4965.60	4962.11	5093.47
	Russian Federation	6621.98	3664.17	3656.05	5848.92	5838.63	5884.57	5604.40	5675.17	5749.30
	Russian Federation (Perm)									
	Serbia				4348.62	4259.18	4205.29	4700.81	4574.95	4760.21
	Shanghai-China									
	Singapore									
	Chinese Taipei							8443.68	8699.00	8768.83
	Thailand	5267.23	2837.90	2902.51	4690.29	5046.52	4935.96	5999.51	5870.23	6084.93
	Trinidad and Tobago									
	Tunisia				4154.39	4668.76	4602.03	4582.28	4535.99	4620.41
	United Arab Emirates									
	Uruguay				5403.32	5743.45	5776.98	4640.34	4555.68	4689.06
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 11.17 Effective sample size 4 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	14115.34	13884.46	14142.52	13993.56	13979.15	13718.71
	Austria	6428.51	6428.00	6511.70	4602.64	4685.38	4483.06
	Belgium	8212.44	8024.14	8054.97	8254.81	8139.11	8262.04
	Canada	22782.00	21516.69	22441.11	20637.21	21109.48	21384.40
	Chile	5466.31	5572.69	5558.84	6715.92	6674.39	6773.64
	Czech Republic	5828.76	5944.47	5962.25	4947.59	4942.37	4736.85
	Denmark	5738.51	5821.92	5489.48	7389.35	7415.55	7441.71
	Estonia	4514.00	4563.53	4456.17	4655.84	4475.01	4588.30
	Finland	5735.82	5794.47	5608.72	8471.69	8511.19	8515.18
	France	4263.26	4196.53	4254.22	4485.79	4168.18	4518.13
	Germany	4881.16	4740.15	4896.67	4941.25	4955.75	4960.55
	Greece	4817.40	4858.44	4659.91	5034.23	4946.07	5016.37
	Hungary	4602.18	4555.09	4567.13	4670.99	4630.69	4682.67
	Iceland	3490.66	3085.72	3529.03	3116.76	3155.83	2364.67
	Ireland	3916.60	3908.96	3800.55	4888.17	4873.34	4727.73
	Israel	5517.37	5697.91	5422.84	4943.03	5013.60	4960.60
	Italy	30037.14	30380.77	29369.02	30596.85	30537.62	29848.67
	Japan	6037.70	6019.35	5993.86	6306.36	6333.06	6189.75
	Korea	4849.14	4961.64	4733.64	4968.00	4979.02	4982.85
	Luxembourg	3099.01	3069.63	3125.74	2115.55	3041.14	2809.01
	Mexico	37516.09	38201.56	36972.85	33266.05	33475.42	33499.88
	Netherlands	4717.50	4734.54	4694.11	4445.97	4277.69	4422.71
	New Zealand	4478.60	4463.01	4550.52	4104.07	4006.20	4237.29
	Norway	4404.22	4501.15	4409.59	4379.21	4643.02	4647.44
	Poland	4777.53	4714.25	4533.31	4520.94	4592.98	4438.34
	Portugal	6248.33	6144.04	6159.07	5659.68	5605.53	5605.99
	Slovak Republic	4434.73	4524.44	4449.79	4543.61	4571.03	4562.21
	Slovenia	5556.90	4992.83	5052.39	3971.57	4509.02	4127.47
	Spain	25701.60	24622.72	25368.31	24734.59	24390.58	24928.07
	Sweden	4478.45	4399.71	4495.04	4692.39	4646.55	4637.94
	Switzerland	11592.61	11080.60	11490.52	11036.39	11012.89	10822.84
	Turkey	4910.45	4966.08	4921.18	4810.83	4786.94	4765.01
	United Kingdom	12022.57	11925.04	11887.13	12491.20	12438.27	12479.57
	United States	5164.00	5189.40	5150.88	4913.79	4735.65	4920.48
	United States (Connecticut)[1]				1636.37	1682.94	1603.76
	United States (Florida)[1]				1890.05	1878.27	1887.04
	United States (Massachusetts)[1]				1680.44	1713.12	1702.05
Partners	Albania	4576.37	4399.33	4526.34	4413.28	4189.60	4648.24
	Argentina	4723.26	4742.81	4759.16	5840.87	5881.39	5840.91
	Azerbaijan	4512.02	4615.03	4462.23			
	Brazil	20000.97	19763.04	19715.44	19118.40	19081.57	18525.93
	Bulgaria	4449.93	4496.29	4488.90	5246.25	5249.49	5255.30
	Colombia	7862.50	7657.69	7793.09	8955.44	9039.23	8906.82
	Costa Rica	4488.99	4471.36	4388.94	4308.80	4501.78	4288.78
	Croatia	4957.75	4846.41	4870.12	4911.34	4997.37	4962.95
	Cyprus[2, 3]				4761.06	4784.86	4385.65
	Georgia	4375.77	4550.65	4532.41			
	Hong Kong-China	4708.95	4773.35	4824.64	4514.42	4541.83	4573.58
	India (Himachal Pradesh)	1541.48	1566.85	1613.43			
	India (Tamil Nadu)	3191.35	3150.27	3175.17			
	Indonesia	5059.35	5061.43	4983.89	5527.47	5506.20	5261.96
	Jordan	6458.91	6441.37	6394.03	6952.33	7009.49	6961.57
	Kazakhstan	5279.36	5219.80	5343.84	5520.17	5675.19	5528.47
	Kyrgyzstan	4946.96	4921.96	4803.96			
	Latvia	4359.95	4449.29	4239.66	4213.69	4207.81	4145.72
	Liechtenstein	289.72	263.33	282.30	212.78	241.94	245.35
	Lithuania	4409.10	4417.71	4426.28	4574.69	4496.80	4435.71
	Macao-China	4201.69	5237.07	3482.67	4563.79	5200.59	4895.25
	Macedonia						
	Malaysia	4960.04	4865.19	4790.22	5074.30	5029.61	4994.67
	Malta	3187.99	3166.13	2574.30			
	Mauritius	3757.57	3665.85	3399.40			
	Moldova	4945.72	5045.15	5047.12			
	Montenegro	4529.66	4441.88	4221.88	3646.09	3735.38	4005.57
	Panama	3866.40	3927.08	3950.96			
	Peru	5899.99	5471.15	5423.52	5926.76	5987.66	5856.21
	Qatar	8963.19	8022.74	5729.17	9204.90	9230.71	10646.02
	Romania	4770.43	4607.29	4727.29	4971.78	4874.85	5024.85
	Russian Federation	5201.79	5266.60	5144.42	5036.46	5158.84	5128.54
	Russian Federation (Perm)				1730.30	1720.69	1697.62
	Serbia	5364.15	5496.13	5473.90	4548.98	4627.73	4557.59
	Shanghai-China	4969.00	5040.60	4852.52	4749.61	4980.26	4845.88
	Singapore	4651.96	3127.70	3907.78	5230.37	5222.79	5043.22
	Chinese Taipei	5686.41	5788.19	5643.54	5879.89	5900.02	5918.20
	Thailand	6118.87	6212.50	6027.70	6471.89	6544.48	6449.58
	Trinidad and Tobago	4617.10	2682.10	3722.81			
	Tunisia	4881.61	4885.57	4861.67	4330.70	4373.59	4300.12
	United Arab Emirates	10593.67	10540.76	10236.14	11277.68	11192.97	11420.66
	Uruguay	5767.32	5486.31	5347.60	5212.18	5254.24	5264.36
	Venezuela (Miranda)	2882.13	2875.09	2831.32			
	Viet Nam				4945.19	4947.83	4932.43

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

- Design Effect 5

11.14

$$Deff_5(r) = \frac{Var_{BRR}(r) + MVar(r)}{Var_{SRS}(r)}$$

shows the inflation of the *total* variance due to the measurement error and due to the complex sampling design. Table 11.18 provides, for each domain and PISA cycle, the design effect 5 values, for each participating country. Table 11.19 provides the corresponding effective sample size.

The product of the first and second design effects equals the product of the third and fourth design effects, and both products are equal to the fifth design effect.

Summary analyses of the design effect

To better understand the evolution of the design effect for a particular country across the PISA cycles, some information related to the design effects and their respective effective sample sizes are presented in Annex C. In particular, the design effect and the effective sample size depend on:

- ***The sample size***, the number of participating schools, the number of participating students and the average within-school sample size, which are provided in Table C.2 (Annex C).
- ***The school variance***, school variance estimates and the intraclass correlation, which are provided respectively in Tables C.3 and C.4 (Annex C).
- ***The stratification variables***, the intraclass correlation coefficient within explicit strata and the percentage of school variance explained by explicit stratification variables, which are provided respectively in Tables C.5 and C.6 (Annex C).

Finally, the standard errors on the mean performance estimates are provided in Table C.1 (Annex C).

Tables 11.20 to 11.26 present the median of the indices presented in Table 11.14 and in Tables C.1 to C.6 in Annex C by cycle and per domain.

[Part 1/2]

Table 11.18 Design effect 5 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	6.20	4.29	3.88	5.98	6.36	5.33	6.86	9.18	7.21
	Austria	3.16	1.94	2.08	6.11	5.66	5.78	7.00	7.15	7.93
	Belgium	7.37	5.10	5.56	4.85	3.81	4.67	6.77	6.87	5.50
	Canada	8.05	4.55	5.15	10.89	12.18	11.57	14.53	11.90	10.53
	Chile	7.78	4.20	3.58				12.24	14.37	11.61
	Czech Republic	3.25	2.55	2.05	8.31	8.63	7.12	8.38	7.03	7.40
	Denmark	2.44	1.88	1.74	4.29	3.81	3.59	5.74	4.31	5.05
	Estonia							5.77	5.67	4.43
	Finland	4.04	1.93	2.23	2.38	2.88	2.60	3.29	3.80	2.62
	France	4.13	2.40	2.50	3.28	3.20	3.13	7.21	5.19	5.16
	Germany	2.49	1.71	1.63	4.49	4.87	4.96	7.59	7.45	7.05
	Greece	12.23	6.91	6.61	7.12	7.99	6.67	7.56	5.03	5.99
	Hungary	8.67	4.69	4.62	3.43	4.39	3.87	5.43	4.51	4.13
	Iceland	0.84	1.33	1.14	0.84	0.83	0.79	1.52	1.60	1.08
	Ireland	4.61	2.25	2.56	3.57	3.20	3.25	6.71	5.28	5.22
	Israel	27.07	12.59	13.15				6.75	7.51	5.07
	Italy	5.06	2.91	2.68	10.63	12.02	9.80	10.83	12.36	9.72
	Japan	19.38	11.67	10.67	6.51	7.51	6.75	7.56	8.02	6.76
	Korea	6.02	2.97	3.07	7.63	6.69	6.75	9.65	8.54	7.19
	Luxembourg	0.89	0.90	1.13	0.87	0.44	0.83	0.75	0.59	0.54
	Mexico	6.85	4.23	4.34	55.44	54.48	48.54	31.66	36.46	35.04
	Netherlands	3.58	2.35	2.38	4.52	4.57	4.07	4.46	4.15	4.00
	New Zealand	2.43	2.07	1.15	2.49	2.38	2.31	3.90	3.16	3.04
	Norway	3.03	2.11	1.91	2.98	2.71	3.11	4.30	3.90	4.92
	Poland	7.28	5.64	5.72	3.94	3.37	3.43	4.31	4.42	3.77
	Portugal	9.91	5.07	5.14	7.46	6.95	6.32	6.63	5.85	5.95
	Slovak Republic				8.36	9.45	9.68	4.00	4.22	3.64
	Slovenia							0.83	0.90	0.85
	Spain	6.35	4.07	3.31	8.01	8.00	7.34	12.39	13.51	15.74
	Sweden	2.52	1.71	1.77	2.97	3.37	3.01	5.44	3.20	2.82
	Switzerland	10.57	6.57	6.70	10.07	9.96	9.89	12.90	12.77	12.36
	Turkey				17.91	20.08	18.29	10.12	13.65	10.52
	United Kingdom	6.07	3.86	3.88	6.55	6.59	5.75	6.44	7.64	6.04
	United States	17.39	12.89	11.13	5.53	5.23	5.00		11.26	8.90
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	5.45	2.31	2.61						
	Argentina	32.83	13.49	13.53				14.46	16.53	15.65
	Azerbaijan							10.24	11.49	12.68
	Brazil	6.33	3.93	3.53	7.54	10.45	8.70	12.40	9.44	9.05
	Bulgaria	10.76	7.00	5.83				15.53	16.54	14.74
	Colombia							9.95	8.27	7.09
	Costa Rica									
	Croatia							5.20	4.20	4.24
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	5.35	2.95	3.05	8.46	9.18	9.18	4.07	3.80	3.38
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	22.31	11.72	11.25	21.15	25.35	23.97	66.74	52.62	71.16
	Jordan							7.86	10.14	6.49
	Kazakhstan									
	Kyrgyzstan							6.85	9.07	7.23
	Latvia	10.36	4.00	7.13	7.62	8.14	8.13	7.98	6.31	5.86
	Liechtenstein	0.57	0.93	0.99	0.53	0.57	0.58	0.57	0.70	0.61
	Lithuania							4.62	5.03	4.45
	Macao-China				1.30	1.37	1.48	0.98	1.14	0.87
	Macedonia	1.93	1.90	1.62						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							0.81	1.15	0.79
	Panama									
	Peru	9.34	4.12	4.86						
	Qatar							0.76	0.79	0.66
	Romania	5.56	3.65	3.42				13.36	12.86	13.71
	Russian Federation	13.69	10.24	8.48	10.63	12.37	10.29	12.48	10.82	9.71
	Russian Federation (Perm)									
	Serbia				8.41	8.66	7.87	6.83	7.02	6.10
	Shanghai-China									
	Singapore									
	Chinese Taipei							14.10	13.95	12.58
	Thailand	9.53	5.62	4.69	6.76	7.01	5.78	6.21	5.09	4.78
	Trinidad and Tobago									
	Tunisia				4.06	4.52	4.06	7.92	8.60	5.98
	United Arab Emirates									
	Uruguay				4.66	6.34	4.11	3.88	3.33	4.10
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 11.18 Design effect 5 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	7.96	10.32	8.86	3.82	4.22	4.44
	Austria	5.71	5.04	6.68	4.30	3.96	4.08
	Belgium	4.53	3.98	4.90	4.19	3.78	4.00
	Canada	6.25	7.89	7.56	9.41	9.25	9.70
	Chile	8.10	8.26	7.30	9.51	9.89	8.72
	Czech Republic	5.95	5.59	5.66	5.58	4.81	5.70
	Denmark	3.65	5.30	4.32	7.15	5.85	6.55
	Estonia	4.74	4.73	4.75	3.05	2.98	2.82
	Finland	3.95	4.01	3.98	5.59	4.57	4.94
	France	4.57	4.04	5.29	3.09	2.93	3.06
	Germany	3.91	4.22	3.85	4.74	4.46	4.85
	Greece	10.25	9.35	9.67	5.61	4.16	6.39
	Hungary	5.70	6.47	6.07	5.68	5.59	5.15
	Iceland	0.79	0.85	0.80	1.18	1.20	1.59
	Ireland	3.84	3.47	4.46	4.40	3.54	3.61
	Israel	6.12	5.71	4.90	9.68	10.07	10.70
	Italy	8.31	12.29	10.41	12.81	14.80	13.46
	Japan	7.26	7.62	7.13	8.80	9.34	9.04
	Korea	9.53	10.10	8.73	10.44	10.75	10.03
	Luxembourg	0.67	0.68	0.64	1.13	0.69	0.84
	Mexico	20.38	20.46	20.50	12.01	11.22	11.53
	Netherlands	16.06	13.51	15.12	6.20	6.41	6.06
	New Zealand	2.43	2.68	2.68	2.21	2.10	1.79
	Norway	3.73	3.67	3.93	4.80	4.28	4.49
	Poland	4.20	5.07	3.79	5.95	7.38	6.02
	Portugal	7.86	6.40	7.59	9.21	9.42	10.19
	Slovak Republic	3.62	4.69	4.46	7.48	5.40	5.95
	Slovenia	0.79	1.02	0.92	1.04	1.07	1.19
	Spain	13.79	14.01	14.25	10.88	11.82	11.36
	Sweden	3.89	4.36	3.39	3.75	2.86	4.30
	Switzerland	8.07	13.05	10.18	9.09	11.67	10.00
	Turkey	9.22	11.30	9.92	11.66	13.63	11.52
	United Kingdom	6.96	9.39	7.91	16.41	15.40	14.54
	United States	7.48	8.10	7.29	8.24	7.97	8.05
	United States (Connecticut)[1]				7.27	6.63	5.77
	United States (Florida)[1]				9.08	8.82	9.12
	United States (Massachusetts)[1]				6.80	6.96	6.59
Partners	Albania	7.51	8.83	9.03	3.61	2.27	2.91
	Argentina	8.75	9.14	9.60	8.77	12.49	12.03
	Azerbaijan	9.12	8.75	7.99			
	Brazil	16.95	17.41	16.87	10.60	11.77	12.92
	Bulgaria	15.69	15.79	13.89	13.64	9.53	11.56
	Colombia	14.80	14.62	15.82	15.40	13.68	14.50
	Costa Rica	7.25	7.91	6.29	10.20	9.11	7.98
	Croatia	5.37	6.13	5.58	7.38	8.01	6.59
	Cyprus[2, 3]				0.58	0.68	0.76
	Georgia	4.04	4.94	4.83			
	Hong Kong-China	3.07	3.96	4.79	5.00	5.21	4.56
	India (Himachal Pradesh)	4.38	5.56	5.23			
	India (Tamil Nadu)	15.75	17.60	13.90			
	Indonesia	16.21	14.38	15.54	17.49	17.98	17.55
	Jordan	8.61	13.06	10.22	10.65	11.35	10.01
	Kazakhstan	6.16	7.18	7.07	7.71	10.49	9.35
	Kyrgyzstan	5.20	6.30	5.18			
	Latvia	6.16	6.76	6.95	3.40	4.87	5.26
	Liechtenstein	0.37	0.71	0.51	0.63	0.50	0.50
	Lithuania	3.46	4.00	5.36	3.80	4.06	4.10
	Macao-China	0.82	0.68	1.09	0.65	0.56	0.62
	Macedonia						
	Malaysia	6.29	6.85	6.25	8.24	7.98	7.57
	Malta	0.61	0.65	0.76			
	Mauritius	0.57	0.55	0.69			
	Moldova	5.23	6.97	6.31			
	Montenegro	1.65	2.76	2.61	0.78	0.77	0.75
	Panama	17.22	16.82	16.12			
	Peru	9.66	11.82	9.15	13.01	11.55	12.63
	Qatar	0.40	0.47	0.67	0.58	0.63	0.54
	Romania	9.88	8.84	8.67	9.85	10.84	8.65
	Russian Federation	7.34	7.98	7.12	5.62	6.46	5.91
	Russian Federation (Perm)				7.18	6.69	6.86
	Serbia	4.66	5.74	4.37	6.46	6.54	7.13
	Shanghai-China	4.57	3.83	4.07	6.60	5.51	7.09
	Singapore	0.62	1.00	0.91	1.02	0.87	1.16
	Chinese Taipei	5.28	6.10	5.39	6.65	4.92	4.76
	Thailand	8.40	10.37	8.71	10.29	11.62	9.72
	Trinidad and Tobago	0.57	0.79	0.63			
	Tunisia	5.66	7.34	5.40	11.57	11.05	8.49
	United Arab Emirates	9.04	7.76	8.27	7.89	8.44	10.31
	Uruguay	4.10	4.77	4.24	5.77	5.16	4.48
	Venezuela (Miranda)	7.96	7.47	7.37			
	Viet Nam				17.46	15.79	15.41

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table 11.19 Effective sample size 5 by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	834.96	666.02	737.57	2097.51	1972.75	2355.92	2067.06	1542.85	1965.53
	Austria	1502.23	1359.95	1283.56	752.13	812.66	795.45	703.61	689.08	621.67
	Belgium	905.14	742.00	669.94	1815.09	2311.05	1883.17	1308.34	1289.80	1609.86
	Canada	3685.55	3625.56	3199.37	2567.92	2295.93	2415.58	1558.20	1903.52	2149.73
	Chile	628.14	648.22	756.65				427.50	364.25	450.89
	Czech Republic	1651.79	1204.27	1495.25	760.65	732.44	887.64	708.22	844.31	801.37
	Denmark	1737.43	1264.11	1351.13	982.35	1107.82	1174.36	789.03	1050.47	897.85
	Estonia							843.80	858.36	1099.28
	Finland	1202.91	1401.53	1214.13	2436.93	2010.96	2226.24	1431.45	1242.11	1800.69
	France	1131.32	1082.32	1035.65	1312.28	1341.86	1372.30	654.26	908.87	914.09
	Germany	2035.92	1655.63	1756.69	1038.79	956.69	939.23	644.43	656.18	693.99
	Greece	382.04	376.77	392.22	649.99	578.79	693.56	644.58	968.04	813.79
	Hungary	563.75	596.57	605.59	1388.19	1086.36	1231.77	827.10	995.12	1085.88
	Iceland	4036.76	1413.78	1633.91	3983.27	4030.53	4241.23	2487.98	2375.01	3501.14
	Ireland	835.82	947.87	832.58	1086.93	1213.33	1195.28	683.50	867.69	877.81
	Israel	166.14	197.21	191.11				679.03	610.64	904.50
	Italy	984.81	950.10	1033.36	1095.37	968.56	1187.82	2010.47	1761.95	2239.07
	Japan	271.14	250.45	273.04	723.05	626.95	697.19	787.30	741.94	880.38
	Korea	828.12	933.77	899.05	713.44	814.05	806.97	536.33	605.95	719.48
	Luxembourg	3970.48	2170.05	1727.06	4509.24	8942.37	4706.38	6112.64	7680.62	8432.43
	Mexico	671.25	606.39	587.19	540.83	550.37	617.66	978.30	849.51	883.93
	Netherlands	699.31	589.18	587.08	883.93	873.90	981.74	1091.55	1174.12	1216.79
	New Zealand	1509.57	987.51	1761.90	1810.59	1897.09	1949.97	1237.01	1524.34	1587.19
	Norway	1368.58	1092.87	1208.39	1362.81	1500.24	1304.94	1092.23	1202.39	953.95
	Poland	501.90	350.31	357.36	1113.62	1298.93	1278.88	1286.43	1255.40	1472.46
	Portugal	462.48	501.78	496.27	617.54	662.64	728.54	770.05	873.35	858.27
	Slovak Republic				878.65	777.53	758.75	1183.12	1121.18	1298.38
	Slovenia							7979.40	7344.39	7802.64
	Spain	978.97	841.43	1045.76	1346.45	1348.72	1469.18	1581.97	1450.70	1245.57
	Sweden	1749.29	1440.60	1379.13	1559.44	1371.11	1534.66	817.25	1387.43	1574.20
	Switzerland	577.25	517.26	507.25	836.16	845.80	851.59	944.77	954.50	986.39
	Turkey				271.07	241.77	265.52	488.14	361.98	469.76
	United Kingdom	1539.83	1344.69	1335.55	1454.81	1445.82	1657.12	2042.08	1722.19	2176.41
	United States	221.20	165.64	191.23	986.82	1043.48	1090.18		498.32	630.35
	United States (Connecticut)[1]									
	United States (Florida)[1]									
	United States (Massachusetts)[1]									
Partners	Albania	913.10	1205.78	1062.51						
	Argentina	121.33	165.31	163.47				300.03	262.46	277.27
	Azerbaijan							506.07	451.02	408.98
	Brazil	773.29	691.94	767.61	590.55	425.99	511.87	749.36	984.13	1026.92
	Bulgaria	432.78	373.61	436.40				289.62	271.96	305.08
	Colombia							449.95	541.63	631.53
	Costa Rica									
	Croatia							1002.28	1241.60	1229.85
	Cyprus[2, 3]									
	Georgia									
	Hong Kong-China	822.65	826.82	798.98	529.37	488.06	487.85	1140.17	1223.86	1374.21
	India (Himachal Pradesh)									
	India (Tamil Nadu)									
	Indonesia	330.19	348.76	362.18	508.69	424.45	448.88	159.52	202.33	149.62
	Jordan							828.64	641.92	1002.85
	Kazakhstan									
	Kyrgyzstan							862.31	650.78	817.16
	Latvia	375.71	536.90	302.70	607.14	568.36	569.02	591.60	747.69	805.63
	Liechtenstein	547.23	188.74	178.07	631.58	578.51	572.56	591.08	485.88	556.99
	Lithuania							1026.20	943.09	1066.30
	Macao-China				962.18	909.52	844.63	4852.87	4186.24	5468.82
	Macedonia	2341.15	1341.27	1558.08						
	Malaysia									
	Malta									
	Mauritius									
	Moldova									
	Montenegro							5467.20	3876.99	5644.78
	Panama									
	Peru	474.35	597.45	506.10						
	Qatar							8231.82	7880.79	9555.56
	Romania	869.17	735.44	788.44				383.13	397.94	373.23
	Russian Federation	489.56	363.17	438.31	561.93	483.10	580.37	464.80	535.74	597.15
	Russian Federation (Perm)									
	Serbia				523.64	508.54	559.41	703.00	682.99	785.96
	Shanghai-China									
	Singapore									
	Chinese Taipei							625.10	631.76	700.61
	Thailand	560.17	526.82	631.55	774.55	747.30	905.64	997.42	1215.92	1296.63
	Trinidad and Tobago									
	Tunisia				1162.55	1044.99	1163.42	585.75	539.23	775.81
	United Arab Emirates									
	Uruguay				1253.14	920.75	1420.62	1245.68	1451.15	1179.17
	Venezuela (Miranda)									
	Viet Nam									

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]
Table 11.19 Effective sample size 5 by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	1790.75	1381.48	1607.84	3793.23	3435.11	3260.66
	Austria	1154.03	1306.57	985.80	1105.41	1201.87	1165.91
	Belgium	1876.64	2134.55	1733.90	2051.08	2273.44	2151.75
	Canada	3710.79	2942.69	3070.74	2289.33	2328.83	2222.05
	Chile	699.63	686.00	776.69	721.28	693.24	786.59
	Czech Republic	1018.48	1084.71	1071.66	954.95	1107.99	934.78
	Denmark	1624.42	1118.25	1371.29	1045.65	1279.88	1142.02
	Estonia	998.24	999.25	995.48	1566.94	1602.09	1692.78
	Finland	1470.33	1449.29	1459.57	1579.55	1932.00	1787.65
	France	940.37	1062.96	811.96	1490.94	1576.33	1508.20
	Germany	1273.30	1180.03	1292.08	1054.05	1120.95	1031.66
	Greece	484.79	531.68	513.91	914.08	1231.36	802.56
	Hungary	807.20	712.00	758.20	846.48	860.10	934.12
	Iceland	4637.23	4288.31	4573.23	2962.51	2917.52	2207.64
	Ireland	1025.20	1133.72	883.17	1141.01	1416.39	1389.64
	Israel	942.00	1008.64	1174.98	522.11	502.22	472.30
	Italy	3720.50	2513.66	2968.50	2426.43	2098.93	2308.67
	Japan	838.72	799.47	853.98	722.01	679.82	702.48
	Korea	523.75	493.77	571.68	482.15	468.23	501.55
	Luxembourg	6848.91	6839.54	7203.88	4642.33	7599.99	6252.64
	Mexico	1876.56	1869.08	1866.22	2814.13	3014.30	2931.14
	Netherlands	296.43	352.31	314.82	719.36	696.25	736.20
	New Zealand	1910.04	1732.62	1732.61	1937.76	2040.81	2393.04
	Norway	1247.74	1268.12	1187.10	977.26	1096.00	1043.27
	Poland	1171.97	969.53	1296.24	774.87	624.27	764.66
	Portugal	801.33	983.50	829.23	621.61	607.33	561.30
	Slovak Republic	1257.66	971.99	1021.73	625.26	866.59	786.77
	Slovenia	7756.78	6052.67	6715.89	5703.72	5528.02	4976.09
	Spain	1876.62	1847.22	1816.10	2326.02	2141.63	2227.30
	Sweden	1173.63	1047.63	1347.06	1264.16	1655.16	1102.34
	Switzerland	1463.01	905.43	1160.25	1234.68	961.81	1122.88
	Turkey	541.70	442.20	503.74	415.69	355.66	420.68
	United Kingdom	1749.14	1297.26	1539.38	771.34	822.22	870.45
	United States	699.26	645.88	717.51	604.32	624.21	618.60
	United States (Connecticut)[1]				233.49	256.02	294.26
	United States (Florida)[1]				208.79	214.93	207.82
	United States (Massachusetts)[1]				253.49	247.48	261.51
Partners	Albania	612.01	520.42	508.83	1314.02	2089.41	1631.21
	Argentina	545.77	522.31	497.15	673.44	472.94	490.99
	Azerbaijan	514.47	536.21	587.33			
	Brazil	1187.67	1155.87	1192.72	1811.46	1631.13	1486.07
	Bulgaria	287.33	285.46	324.52	387.37	554.47	456.90
	Colombia	535.28	541.77	500.58	589.09	663.40	625.65
	Costa Rica	631.77	578.82	727.63	451.19	505.32	576.42
	Croatia	930.74	814.30	895.29	678.72	625.06	759.87
	Cyprus[2, 3]				8827.29	7512.26	6670.59
	Georgia	1149.60	939.95	961.94			
	Hong Kong-China	1575.00	1220.65	1008.91	934.51	896.65	1023.20
	India (Himachal Pradesh)	368.64	290.78	309.01			
	India (Tamil Nadu)	203.79	182.39	230.95			
	Indonesia	316.75	357.18	330.55	321.36	312.66	320.28
	Jordan	753.46	496.48	634.43	660.66	619.85	702.80
	Kazakhstan	877.89	754.13	764.96	753.67	553.54	620.93
	Kyrgyzstan	958.13	790.86	963.09			
	Latvia	731.39	666.13	647.85	1266.28	884.64	818.40
	Liechtenstein	880.73	466.57	650.72	461.99	583.35	583.28
	Lithuania	1307.31	1133.24	844.08	1214.96	1138.76	1127.25
	Macao-China	7289.19	8689.55	5444.57	8248.62	9595.98	8618.87
	Macedonia						
	Malaysia	795.22	729.56	800.34	631.00	651.00	686.81
	Malta	5662.79	5286.74	4535.68			
	Mauritius	8120.42	8391.53	6758.55			
	Moldova	993.29	745.56	823.49			
	Montenegro	2928.94	1745.73	1848.18	6066.96	6150.89	6295.75
	Panama	230.47	236.00	246.26			
	Peru	619.38	506.35	653.77	463.83	522.74	478.00
	Qatar	22777.39	19415.46	13538.92	18759.99	17460.01	20376.04
	Romania	483.55	540.40	550.80	515.26	468.01	586.57
	Russian Federation	723.28	665.29	745.99	930.87	809.86	885.04
	Russian Federation (Perm)				245.37	263.12	256.66
	Serbia	1185.36	962.50	1263.40	724.75	716.35	656.55
	Shanghai-China	1119.02	1334.04	1257.51	784.75	939.62	730.25
	Singapore	8519.47	5257.65	5827.97	5421.38	6345.69	4785.86
	Chinese Taipei	1104.96	955.60	1082.08	909.31	1229.42	1269.85
	Thailand	740.86	600.27	714.95	641.80	568.52	679.85
	Trinidad and Tobago	8354.94	6030.73	7641.20			
	Tunisia	874.96	675.28	917.84	380.84	398.94	518.85
	United Arab Emirates	1202.69	1400.60	1314.29	1457.01	1361.92	1115.50
	Uruguay	1453.88	1248.10	1406.50	920.43	1030.17	1187.45
	Venezuela (Miranda)	364.44	388.43	393.66			
	Viet Nam				284.00	314.07	321.87

1. Only public.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 11.20 Median of the design effect 3 per cycle and per domain across the 35 countries that participated in every cycle

	Reading	Mathematics	Science
PISA 2000	5.90	3.68	2.93
PISA 2003	6.02	6.25	5.45
PISA 2006	6.69	6.26	5.63
PISA 2009	5.96	6.40	6.61
PISA 2012	5.51	5.38	5.79

In PISA 2000, student performance estimates for a particular domain were only provided for students who responded to testing material from that domain, while for PISA 2003 onwards student proficiency estimates were provided for all domains. For PISA 2000 about five-ninths of the students were assessed in the minor domains (Adams and Wu, 2002). This difference explains why the design effects in mathematics and science for PISA 2000 are so low in comparison with all other design effects.

Table 11.21 presents summary information about the standard errors of national mean achievement across PISA cycles.

Table 11.21 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

	Reading	Mathematics	Science
PISA 2000	3.10	3.26	3.18
PISA 2003	2.88	3.00	3.08
PISA 2006	3.18	2.89	2.79
PISA 2009	2.66	2.83	2.80
PISA 2012	2.82	2.73	2.75

The standard errors, on average, have decreased between the PISA 2000 and PISA 2012 data collection for mathematics and science, though not for reading. This decrease is associated with the continuously increasing school sample size. Note that, generally speaking, the sample size increase in a given country, in PISA 2009 and PISA 2012 compared with earlier cycles, was intended to provide adequate data for regional or other subgroup estimates. Consequently the reduction in standard error for the national mean achievement was often not particularly great for countries with a noticeable increase in sample size. In other words, the sample size increased, but so did the design effects for the participating countries mean achievement estimates.

This reduction of the standard errors might also be explained by a better efficiency of the explicit stratification variables. Although as can be found in Table 11.23 the median percentage of school variance explained by explicit stratification variables has not consistently risen or fallen over the five cycles.

Table 11.22 shows that median school sample sizes have generally been increasing across PISA cycles from 174 schools in PISA 2000 to 209 schools in PISA 2012. There is a jump between 2009 and 2012, which could be in part due to some countries using a lower target cluster size (TCS) and thus sampling more schools. For example, Australia's TCS is 25 in 2012 as against 50 in 2009.

Table 11.22 Median of the number of participating schools for each PISA cycle for the 35 countries that participated in every cycle

	Number of schools
PISA 2000	174
PISA 2003	193
PISA 2006	190
PISA 2009	193
PISA 2012	209

Table 11.23 shows information about the size of the between-school variance across PISA cycles.

Table 11.23 **Median of the school variance estimate for each domain and PISA cycle for the 35 countries that participated in every cycle**

	Reading	Mathematics	Science
PISA 2000	3305	3127	2574
PISA 2003	2481	2620	2270
PISA 2006	2982	2746	2502
PISA 2009	2256	2481	2266
PISA 2012	2736	2753	2522

To understand the pattern of school variance estimates, it is important to recall how the school membership was implemented in the conditioning model. In PISA 2000 and PISA 2003, the conditioning variable consists of the truncated school average of student performance weighted maximum likelihood estimates in the major domain. In PISA 2006 to PISA 2012, the conditioning variables consist of *n*-1 dummy variables; with *n* being the corrected number of participating schools, where non UH schools with less than 8 students are grouped together into one school and UH schools, if present are grouped into another, where UH refers to the easier one hour ("Une Heure") booklet. The method used in the first two PISA studies seemed to generate an underestimation of the school variance estimates in the minor domains. This bias might therefore explain why the largest school variance estimate in PISA 2000 and in PISA 2003 was associated with the major domain, respectively reading literacy and mathematic literacy.

Table 11.24 **Median of the intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle**

	Reading	Mathematics	Science
PISA 2000	0.37	0.36	0.33
PISA 2003	0.30	0.34	0.28
PISA 2006	0.38	0.36	0.35
PISA 2009	0.33	0.33	0.34
PISA 2012	0.35	0.34	0.34

Table 11.25 **Median of the within explicit strata intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle**

	Reading	Mathematics	Science
PISA 2000	0.25	0.22	0.23
PISA 2003	0.20	0.23	0.19
PISA 2006	0.26	0.23	0.20
PISA 2009	0.20	0.22	0.22
PISA 2012	0.23	0.23	0.23

Table 11.26 **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**

	Reading	Mathematics	Science
PISA 2000	20.1	17.9	18.8
PISA 2003	22.5	21.6	20.5
PISA 2006	33.7	25.6	29.9
PISA 2009	31.2	27.6	30.8
PISA 2012	34.1	30.1	29.7

Sampling for the problem solving component

Out of the 65 countries and economies that participated in PISA 2012, 44 also implemented the computer-based assessment of problem solving. Of these, 12 countries and economies only assessed problem solving, while 32 also assessed mathematics and (digital) reading on computers. In all 44 countries/economies, only a random sub-sample of students who participated in the paper-based assessment of mathematics were sampled to be administered the assessment of problem solving. However, as long as at least one student in a participating school was sampled for the computer-based assessment, all students in the PISA sample from that school received multiple imputations (plausible values) of performance in problem solving. This is similar to the procedure used to impute plausible values for minor domains in PISA.

Table 11.27 Sample size for performance in mathematics and problem solving

		Mathematics		Problem solving			
		Number of schools with valid data (unweighted)	Number of students with valid data (unweighted)	Number of schools with valid data (unweighted)	Number of students with valid data (unweighted)	Number of students with valid data sampled for the assessment of problem solving (unweighted)	Number of students who were administered the assessment of problem solving (unweighted)
		(1)	(2)	(3)	(4)	(5)	(6)
OECD	Australia	775	14481	775	14481	5922	5612
	Austria	191	4755	191	4755	1376	1331
	Belgium	287	8597	287	8597	2309	2147
	Canada	885	21544	885	21544	5415	4602
	Chile	221	6856	221	6856	1674	1578
	Czech Republic	297	5327	297	5327	3229	3076
	Denmark	341	7481	341	7481	2104	1948
	Estonia	206	4779	206	4779	1412	1367
	Finland	311	8829	311	8829	3685	3531
	France	226	4613	226	4613	1509	1345
	Germany	230	5001	230	5001	1426	1350
	Greece	188	5125	0	0	0	0
	Hungary	204	4810	204	4810	1355	1300
	Iceland	134	3508	0	0	0	0
	Ireland	183	5016	183	5016	1303	1190
	Israel	172	5055	172	5055	1445	1346
	Italy	1194	31073	208	5495	1554	1371
	Japan	191	6351	191	6351	3178	3014
	Korea	156	5033	156	5033	1351	1336
	Luxembourg	42	5258	0	0	0	0
	Mexico	1471	33806	0	0	0	0
	Netherlands	179	4460	179	4460	2258	1752
	New Zealand	177	4291	0	0	0	0
	Norway	197	4686	197	4686	1463	1240
	Poland	184	4607	184	4607	1256	1227
	Portugal	195	5722	195	5722	1631	1446
	Slovak Republic	231	4678	231	4678	1589	1465
	Slovenia	338	5911	338	5911	2179	2065
	Spain	902	25313	368	10175	2866	2709
	Sweden	209	4736	209	4736	1337	1258
	Switzerland	411	11229	0	0	0	0
	Turkey	170	4848	170	4848	2022	1995
	United Kingdom[1]	507	12659	170	4185	1963	1458
	United States	162	4978	162	4978	1300	1273
Partners	Albania	204	4743	0	0	0	0
	Argentina	226	5908	0	0	0	0
	Brazil	839	19204	241	5506	1590	1463
	Bulgaria	188	5282	188	5282	2333	2145
	Colombia	352	9073	352	9073	2595	2307
	Costa Rica	193	4602	0	0	0	0
	Croatia	163	5008	163	5008	2016	1924
	Cyprus[2, 3]	117	5078	117	5078	2630	2503
	Hong Kong-China	148	4670	148	4670	1367	1325
	Indonesia	209	5622	0	0	0	0
	Jordan	233	7038	0	0	0	0
	Kazakhstan	218	5808	0	0	0	0
	Latvia	211	4306	0	0	0	0
	Liechtenstein	12	293	0	0	0	0
	Lithuania	216	4618	0	0	0	0
	Macao-China	45	5335	45	5335	1577	1565
	Malaysia	164	5197	164	5197	2072	1929
	Montenegro	51	4744	51	4744	2101	1845
	Peru	240	6035	0	0	0	0
	Qatar	157	10966	0	0	0	0
	Romania	178	5074	0	0	0	0
	Russian Federation	227	5231	227	5231	1574	1543
	Serbia	153	4684	153	4684	1930	1777
	Shanghai-China	155	5177	155	5177	1213	1203
	Singapore	172	5546	172	5546	1438	1394
	Chinese Taipei	163	6046	163	6046	1512	1484
	Thailand	239	6606	0	0	0	0
	Tunisia	153	4407	0	0	0	0
	United Arab Emirates	458	11500	458	11500	3418	3262
	Uruguay	180	5315	180	5315	2048	2013
	Viet Nam	162	4959	0	0	0	0

1. Only England took part in the assessment of problem solving.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 11.28 Sample size for performance in mathematics and problem solving, by adjudicated regions

		Mathematics		Problem solving			
		Number of schools with valid data (unweighted) (1)	Number of students with valid data (unweighted) (2)	Number of schools with valid data (unweighted) (3)	Number of students with valid data (unweighted) (4)	Number of students with valid data sampled for the assessment of problem solving (unweighted) (5)	Number of students who were administered the assessment of problem solving (unweighted) (6)
OECD	Belgium (Flemish community)	174	4877	174	4877	1465	1334
	Spain (Andalusia)	52	1434	33	910	244	232
	Spain (Aragon)	51	1393	6	159	45	44
	Spain (Asturias)	56	1611	4	120	31	28
	Spain (Balearic Islands)	54	1435	4	100	31	24
	Spain (Basque Country)	174	4739	174	4739	1395	1351
	Spain (Cantabria)	54	1523	4	111	32	32
	Spain (Castile and Leon)	55	1592	7	201	54	52
	Spain (Catalonia)	51	1435	51	1435	392	354
	Spain (Extremadura)	53	1536	5	150	42	39
	Spain (Galicia)	56	1542	8	202	59	54
	Spain (La Rioja)	54	1532	4	85	29	27
	Spain (Madrid)	51	1542	20	592	157	144
	Spain (Murcia)	52	1374	6	141	37	36
	Spain (Navarre)	51	1530	4	135	32	31
	United Kingdom (Scotland)	111	2945	0	0	0	0
	United States (Connecticut)[1]	50	1697	0	0	0	0
	United States (Florida)[1]	54	1896	0	0	0	0
	United States (Massachusetts)[1]	49	1723	0	0	0	0
Partners	Argentina (Ciudad Autónoma de Buenos Aires)	49	1336	0	0	0	0
	Russian Federation (Perm Territory)	63	1761	0	0	0	0
	United Arab Emirates (Abu Dhabi)	119	3163	119	3163	920	882
	United Arab Emirates (Dubai)	209	4974	209	4974	1523	1461

1. Only public.

Tables 11.27 and 11.28 (by adjudicated region) compare the final samples (after school replacement) for mathematics and problem solving.

- Column 1 shows the overall number of schools with valid data for mathematics in the PISA 2012 database.
- Column 2 shows the students with valid data in mathematics. This is the number of students with data included in the main database. All these students have imputed values for performance in mathematics, reading and science. Students are considered as participating in the assessment of mathematics if they were sampled to sit the paper-based assessment (all booklets included mathematics questions) and attended a test session. Those who only attended the questionnaire session but provided at least a description of their father's or mother's occupation are also regarded as participants.
- Column 3 shows the number of schools with valid data for problem solving in the PISA 2012 computer-based assessments database.
- Column 4 shows the number of students with valid data in problem solving. This corresponds to all participating students (Column 2) within schools who were sampled for problem solving in PISA 2012 and were included in the database (Column 3). For all these students, performance in problem solving could be imputed. All these students contributed to the statistics presented in this publication (with the exception of statistics based on item-level performance).
- Column 5 shows the number of students included in the database who were sampled for the assessment of problem solving. These are the students with valid data who were sampled to sit the computer-based assessment and assigned a form (the computer equivalent of a paper booklet) containing at least one cluster of problem-solving questions.
- Column 6 shows the number of students who were actually assessed in problem solving. These are the students sampled for the assessment of problem solving who actually attended the computer-based assessment session and were administered the test. All these students contributed to statistics based on item-level performance in this volume. Differences between the number of students in Columns 5 and 6 can occur for several reasons: students who skipped the computer-based session; students who did not reach any of the problem-solving questions in their test form; and technical problems with the computer.

Note

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

References

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

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

OECD (2014), *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science* (Volume I, Revised edition, February 2014), PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201118-en

12

Scaling Outcomes

This chapter describes the outcomes of applying Item Response Theory (IRT) scaling and plausible value methodology to the PISA 2012 assessment data.

INTERNATIONAL CHARACTERISTICS OF THE ITEM POOL

When Main Survey data were received from each participating country, they were first verified and cleaned using the procedures outlined in Chapter 10. Files containing the achievement data were prepared and national-level Rasch and traditional test analyses were undertaken. The results of these analyses were included in the reports that were returned to each participant (see Chapter 9).

After processing at the national level, a set of international-level analyses was undertaken. Some involved summarising national analyses, while others required an analysis of the international data set.

The final international cognitive data set (that is, the data set of coded achievement booklet responses – available as *INT_COG12_Dec03.txt*[1]) consisted of 485 490 students from the participating countries and economies. Table 12.1 shows the total number of students included in the PISA 2012 database, broken down by participating country and test booklet. Countries that implemented the easier set of booklets (see Chapter 2) are shown in italics in this table. For countries and economies that participated in financial literacy, there is variation in the number of students that responded to each booklet because booklet assignment algorithm, while ensuring random assignment, did not ensure a uniform distribution. Weight adjustments were applied to compensate for this non-uniform, but random, assignment of booklets to students.

[Part 1/2]

Table 12.1 **Number of sampled students by country/economy and paper-based booklet**

		Booklets														
		1 or 21	2 or 22	3 or 23	4 or 24	5 or 25	6 or 26	7 or 27	8	9	10	11	12	13	UH[1]	Total
OECD	**Australia**	1 145	1 148	1 119	1 101	1 164	1 152	1 194	1 202	1 165	1 025	1 020	1 024	1 022		14 481
	Austria	369	364	367	365	349	357	362	373	360	355	361	366	371	36	4 755
	Belgium	697	674	699	677	689	686	687	667	678	549	556	550	571	217	8 597
	Canada	1 656	1 662	1 665	1 665	1 667	1 627	1 623	1 689	1 665	1 675	1 660	1 642	1 648		21 544
	Chile	552	529	553	522	511	525	528	536	521	507	539	520	513		6 856
	Czech Republic	426	428	406	419	424	425	416	430	429	359	358	355	361	91	5 327
	Denmark	567	572	585	570	569	581	564	564	552	593	551	565	566	82	7 481
	Estonia	413	392	380	403	412	410	403	405	402	294	284	287	294		4 779
	Finland	683	689	669	691	667	680	667	671	659	668	672	686	679	48	8 829
	France	380	372	376	355	363	379	361	362	362	306	326	334	337		4 613
	Germany	379	385	353	351	374	361	386	390	387	367	377	370	382	139	5 001
	Greece	387	399	387	388	404	396	388	392	385	406	405	396	392		5 125
	Hungary	366	353	377	382	377	368	356	363	364	377	384	372	371		4 810
	Iceland	262	272	277	273	262	267	279	262	263	290	273	267	261		3 508
	Ireland	401	382	380	390	381	384	393	378	373	388	382	385	399		5 016
	Israel	381	372	389	385	371	580	381	563	388	455	263	269	258		5 055
	Italy	2 619	2 604	2 612	2 650	2 634	2 663	2 626	2 651	2 629	1 844	1 846	1 861	1 834		31 073
	Japan	482	487	486	490	502	481	494	481	475	495	490	493	495		6 351
	Korea	385	377	382	386	384	388	379	394	396	392	391	394	385		5 033
	Luxembourg	407	399	403	405	406	402	407	403	406	401	405	409	405		5 258
	Mexico	2 604	2 607	2 605	2 647	2 638	2 607	2 621	2 597	2 579	2 561	2 604	2 586	2 550		33 806
	Netherlands	341	332	330	338	337	340	333	324	335	338	343	319	323	127	4 460
	New Zealand	342	340	345	357	332	348	345	347	351	288	280	306	310		4 291
	Norway	363	375	366	365	360	363	368	354	355	352	360	359	346		4 686
	Poland	388	402	387	375	396	393	384	383	392	287	284	265	271		4 607
	Portugal	438	430	460	438	441	447	450	452	444	416	427	438	441		5 722
	Slovak Republic	372	378	389	355	391	378	387	398	393	294	310	296	302	35	4 678
	Slovenia	473	466	449	464	432	451	466	455	450	408	407	421	424	145	5 911
	Spain	1 985	1 988	2 000	1 995	1 974	1 986	2 001	1 976	1 998	1 873	1 836	1 840	1 861		25 313
	Sweden	379	371	377	368	374	366	365	361	337	343	355	367	373		4 736
	Switzerland	877	867	870	850	863	855	871	866	874	848	844	872	872		11 229
	Turkey	372	378	375	379	375	362	372	382	377	370	374	364	368		4 848
	United Kingdom	972	958	987	977	993	989	954	955	940	959	985	977	1 010	3	12 659
	United States	397	421	422	419	438	439	429	415	423	296	291	289	299		4 978
	United States (Connecticut)	135	128	135	137	126	136	123	125	133	127	132	121	139		1 697
	United States (Florida)	148	142	149	144	144	144	146	155	146	145	144	150	139		1 896
	United States (Massachusetts)	126	136	143	138	135	132	128	132	131	128	128	133	133		1 723

Note: Countries which implemented the easier set of booklets are shown in italics.

1. This refers to one-hour booklet, see Chapter 2 for more details.

2. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

3. Footnote by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

4. Cyprus data is not included in the international database.

[Part 2/2]

Table 12.1 Number of sampled students by country/economy and paper-based booklet

		Booklets													
		1 or 21	2 or 22	3 or 23	4 or 24	5 or 25	6 or 26	7 or 27	8	9	10	11	12	13	UH[1]	Total
Partners	**Albania**	368	339	360	383	378	376	367	369	359	367	362	356	359		4 743
	Argentina	455	438	451	474	467	461	464	452	456	438	456	453	443		5 908
	Brazil	1 455	1 503	1 463	1 490	1 500	1 460	1 467	1 467	1 490	1 477	1 477	1 491	1 464		19 204
	Bulgaria	419	407	410	391	401	407	418	411	408	404	402	404	400		5 282
	Colombia	753	775	766	775	766	785	761	758	767	549	536	554	528		9 073
	Costa Rica	344	346	337	335	328	334	331	342	340	346	336	334	338	211	4 602
	Croatia	416	426	433	425	421	426	421	434	439	292	301	291	283		5 008
	Cyprus[2, 3]	391	390	400	390	370	387	387	397	389	396	387	394	400		5 078
	Hong Kong-China	369	364	352	361	348	358	372	360	359	360	348	354	365		4 670
	Indonesia	424	425	417	431	422	431	434	441	432	449	445	438	433		5 622
	Jordan	554	537	535	551	539	541	541	538	537	540	542	549	534		7 038
	Kazakhstan	455	456	453	449	436	433	434	443	442	452	450	452	453		5 808
	Latvia	357	345	349	351	361	360	351	360	361	284	282	270	275		4 306
	Liechtenstein	21	24	23	22	24	25	22	22	23	20	25	22	20		293
	Lithuania	351	343	337	356	352	352	354	361	359	361	370	356	366		4 618
	Macao-China	410	410	413	410	411	409	413	410	411	406	411	411	410		5 335
	Malaysia	401	395	409	409	403	396	393	401	398	390	398	407	397		5 197
	Montenegro	369	371	376	373	366	354	365	356	368	369	369	353	355		4 744
	Peru	466	466	464	452	455	468	459	472	467	467	462	471	466		6 035
	Qatar	850	860	837	843	829	822	839	842	842	859	836	850	857		10 966
	Romania	384	391	391	399	398	389	394	394	394	386	390	381	383		5 074
	Russian Federation	440	430	434	428	440	439	449	442	448	326	319	316	320		5 231
	Russian Federation (Perm)	133	126	134	137	137	136	135	136	138	141	137	134	137		1 761
	Serbia	350	359	371	363	367	366	362	355	355	361	357	359	359		4 684
	Shanghai-China	442	439	439	442	444	446	443	440	447	300	297	299	299		5 177
	Singapore	427	415	420	425	419	432	439	423	422	434	434	436	420		5 546
	Chinese Taipei	452	468	473	465	473	468	470	468	462	461	456	466	464		6 046
	Thailand	510	513	501	503	494	490	507	500	501	521	524	525	517		6 606
	Tunisia	338	338	338	346	346	350	339	336	353	323	339	332	329		4 407
	United Arab Emirates	895	893	884	891	899	879	886	880	886	890	869	870	878		11 500
	Uruguay	410	412	416	408	408	399	385	411	406	408	418	418	416		5 315
	Viet Nam	393	374	388	377	377	376	372	377	378	385	386	383	393		4 959
	Total number of students[4]	**38 080**	**37 967**	**38 028**	**38 079**	**38 068**	**38 216**	**38 024**	**38 254**	**37 965**	**35 545**	**35 381**	**35 383**	**35 366**	**1 134**	**485 490**

Note: Countries which implemented the easier set of booklets are shown in italics.

[1] This refers to one-hour booklet, see Chapter 2 for more details.

[2] Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[3] Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

[4] Cyprus data is not included in the international database.

Forty-four countries and economies participated in the PISA 2012 computer-based assessment of problem solving, in addition to the paper-based assessment. An optional computer-based assessment of reading (referred to as the digital reading assessment, or DRA) and mathematics (referred to as the computer-based assessment of mathematics, or CBAM) was administered in 32 countries and economies. In this report, these computer-based assessments will be referred to collectively as CBA. The final international computer-based assessment cognitive data set (that is, the data set of coded achievement computer form responses – available as *CBA_COG12_MAR31.txt*[2]) consisted of 276 401 students from 44 participating countries/economies. The number of cases included in the CBA cognitive data set is the same as in the final international cognitive data set for all participating countries and economies except for Brazil, Italy and Spain, which chose to have schools sub-sampled for CBA from their large national school sample (see Chapter 4 for details of CBA sampling) and for the United Kingdom where problem solving was administered in England only.

Tables 12.2a and 12.2b show the total number of students included in the PISA 2012 CBA database, broken down by participating country/economy and CBA test form. For the students who were not sampled for CBA, the test form code is 97.

[Part 1/2]

Table 12.2a Number of sampled students by country/economy and computer-based form (problem solving, mathematics and reading)

		Forms													
		41	42	43	44	45	46	47	48	49	50	51	52	53	54
OECD	Australia	483	497	489	498	472	485	490	501	490	501	496	514	496	486
	Austria	111	118	115	113	104	108	108	118	120	118	127	116	118	112
	Belgium	192	193	188	189	190	178	175	176	174	183	184	195	185	189
	Canada	443	461	446	423	414	422	413	427	440	434	469	451	469	457
	Chile	131	138	138	128	143	135	134	128	141	127	133	138	144	137
	Denmark	171	180	166	170	169	172	168	176	184	184	175	179	172	188
	Estonia	123	128	120	114	113	103	107	102	107	108	106	120	119	110
	France	128	118	126	127	129	122	116	130	130	137	132	132	128	130
	Germany	110	118	118	116	119	126	128	127	120	130	123	121	116	123
	Hungary	117	114	114	111	114	117	123	123	115	126	113	116	102	108
	Ireland	102	97	103	102	106	103	112	109	109	118	107	124	121	120
	Israel	128	93	130	99	102	90	99	99	140	98	138	96	105	139
	Italy	129	133	131	139	128	131	131	129	131	132	136	125	133	138
	Japan	271	274	273	267	262	260	254	250	248	254	258	256	255	256
	Korea	119	117	118	122	114	112	109	115	110	100	111	114	108	106
	Norway	122	116	115	132	126	124	123	126	131	130	121	123	120	113
	Poland	95	109	108	108	111	111	114	118	117	116	99	109	118	100
	Portugal	143	139	144	137	142	137	138	141	124	136	133	147	135	142
	Slovak Republic	132	121	120	130	130	129	136	135	137	125	134	133	124	128
	Slovenia	191	199	190	186	186	192	182	191	190	184	178	174	187	185
	Spain	247	246	233	243	249	244	232	238	225	226	237	229	220	237
	Sweden	114	117	108	109	109	103	103	100	102	98	105	105	108	111
	United States	115	113	105	101	100	105	99	104	103	110	112	112	114	110
Partners	Brazil	137	147	129	129	127	127	128	137	135	130	129	137	139	135
	Colombia	215	216	220	221	223	221	207	217	218	214	231	216	222	214
	Hong Kong-China	117	108	113	112	110	107	111	114	123	124	118	117	119	117
	Macao-China	136	135	135	133	128	128	133	134	136	127	129	128	127	130
	Russian Federation	127	134	139	140	143	141	138	140	135	136	138	137	132	131
	Shanghai-China	99	102	105	109	102	99	96	93	98	94	93	97	102	110
	Singapore	123	121	124	114	117	117	115	122	119	120	118	119	126	118
	Chinese Taipei	138	138	136	132	131	128	125	127	123	119	109	120	121	128
	United Arab Emirates	260	271	267	289	288	296	277	281	296	290	300	286	281	297

[Part 2/2]

Table 12.2a Number of sampled students by country/economy and computer-based form (problem solving, mathematics and reading)

		Forms										Total sampled students	Total not-sampled students
		55	56	57	58	59	60	61	62	63	64		
OECD	Australia	487	493	494	487	478	491	490	499	511	506	11 834	2 647
	Austria	111	118	117	119	114	109	106	102	111	118	2 731	2 024
	Belgium	200	196	195	192	204	207	209	210	210	203	4 617	3 980
	Canada	464	461	464	475	459	466	477	471	456	455	10 817	10 727
	Chile	135	149	145	145	145	143	148	146	144	146	3 341	3 515
	Denmark	181	173	180	170	176	161	159	171	159	165	4 149	3 332
	Estonia	108	118	119	125	127	133	140	136	122	129	2 837	1 942
	France	138	125	126	123	109	118	117	119	123	129	3 012	1 601
	Germany	125	115	121	119	115	112	123	121	120	115	2 881	2 120
	Hungary	105	102	106	111	114	116	125	121	118	115	2 746	2 064
	Ireland	127	112	108	110	110	108	95	102	104	104	2 613	2 403
	Israel	106	130	106	100	145	101	101	142	98	92	2 677	2 378
	Italy	142	128	130	125	126	115	113	118	123	123	3 089	2 406
	Japan	262	265	262	263	275	272	277	277	277	283	6 351	0
	Korea	112	117	114	107	105	103	110	108	111	113	2 675	2 358
	Norway	117	124	116	116	117	123	128	117	121	123	2 924	1 762
	Poland	102	93	101	98	97	104	111	112	102	114	2 567	2 040
	Portugal	134	133	134	132	130	137	134	131	129	140	3 272	2 450
	Slovak Republic	130	133	136	130	137	143	131	125	131	135	3 145	1 533
	Slovenia	170	155	165	173	185	173	180	188	191	190	4 385	1 526
	Spain	248	244	222	233	240	255	248	242	257	256	5 751	4 424
	Sweden	112	106	121	115	128	123	121	119	120	114	2 671	2 065
	United States	105	103	106	99	111	106	107	108	110	114	2 572	2 406
Partners	Brazil	129	135	139	125	119	134	127	132	130	136	3 172	2 334
	Colombia	219	220	226	226	212	212	192	202	210	199	5 173	3 900
	Hong Kong-China	110	106	106	102	109	112	110	117	114	118	2 714	1 956
	Macao-China	131	129	133	135	133	128	129	130	128	132	3 147	2 188
	Russian Federation	130	125	123	129	130	135	129	125	127	122	3 186	2 045
	Shanghai-China	111	112	108	104	103	97	96	95	94	90	2 409	2 768
	Singapore	123	121	121	113	121	115	118	122	121	125	2 873	2 673
	Chinese Taipei	124	119	127	131	126	134	136	128	129	134	3 063	2 983
	United Arab Emirates	298	298	287	279	285	267	258	263	257	261	6 732	4 768

Table 12.2b Number of sampled students by country/economy and computer-based form (problem solving only)

		Forms								Total sampled students	Total not-sampled students
		31	32	33	34	35	36	37	38		
OECD	**Czech Republic**	405	403	406	402	400	395	408	410	3 229	2 098
	Finland	461	458	458	468	474	470	450	446	3 685	5 144
	Netherlands	284	285	284	271	289	285	272	288	2 258	2 202
	Turkey	256	250	242	249	258	258	257	252	2 022	2 826
	United Kingdom[1]	249	248	234	255	241	248	240	248	1 963	2 222
Partners	**Bulgaria**	302	286	287	298	294	293	282	291	2 333	2 949
	Croatia	251	256	240	253	248	258	256	254	2 016	2 992
	Cyprus[2, 3]	337	328	316	335	330	318	338	328	2 630	2 448
	Malaysia	266	260	255	266	263	250	254	258	2 072	3 125
	Montenegro	268	264	261	267	269	259	257	256	2 101	2 643
	Serbia	242	243	248	241	242	233	241	240	1 930	2 754
	Uruguay	261	257	260	252	240	259	269	250	2 048	3 267

1. England only.
2. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
3. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Generally the CBA Target Cluster Size (ETCS) has been set to 14 for countries that participated only in the problem solving assessment and to 18 for countries that participated in CBA option (Chapter 4). But in a few countries there was a variation in the ETCS. Table 12.3 shows the corresponding Target Cluster Size (TCS) and ETCS for countries and economies that participated in the CBA and problem solving assessments.

Table 12.3 [Part 1/2] ETCS and TCS for countries/economies that participated in CBA and problem solving

		CBA and PS	CBA Target Cluster Size (ETCS)	Target Cluster Size (TCS)
OECD	**Australia**	PS+CBA	20	25
	Austria	PS+CBA	18, but 0 in SEN[1] schools	35
	Belgium	PS+CBA	20 for BFL non-SEN schools; 18 for BGR and BFR non-SEN schools; 0 otherwise	43 for BFL non-SEN schools; 35 otherwise
	Canada	PS+CBA	18	100 for strata 3 and 37; 95 for stratum 43, all students for strata 10, 26, 66, 76, 86, and 96; 35 for all other strata
	Chile	PS+CBA	18	40
	Czech Republic	PS	14 (0 in stratum 80, 81, and 82)	43
	Denmark	PS+CBA	15, but 0 in SEN schools, and all students in stratum 5 schools (Faroes)	28 for strata 1, and 4, 20 for strata 2, and 3, all for stratum 5
	Estonia	PS+CBA	18	43
	Finland	PS	14	20 for Stratum '01', '02', '05', '08', '11', and 35 elsewhere
	France	PS+CBA	16	32
	Germany	PS+CBA	14, but 0 in SEN schools	25
	Hungary	PS+CBA	18	35
	Ireland	PS+CBA	18	35
	Israel	PS+CBA	18	43
	Italy	PS+CBA	18 (only in the subsample of schools)	All students in all Valle D'Aosta strata; 43 otherwise
	Japan	PS+CBA	35	35
	Korea	PS+CBA	18	35
	Netherlands	PS	15	30
	Norway	PS+CBA	18	30
	Poland	PS+CBA	18	43
	Portugal	PS+CBA	21	40
	Slovak Republic	PS+CBA	18	45
	Slovenia	PS+CBA	18	37 for stratum 1, 31 for all other strata
	Spain	PS+CBA	18 for national subsample. In the oversample: 18 for regions 9 and 16 and 0 otherwise	43 for national subsample. In oversample: 48 for adjudicated region 12; 140 for adjudicated region 18; 35 for all other regions
	Sweden	PS+CBA	16	30
	Turkey	PS	14	35
	United Kingdom[2]	PS	14	30
	United States	PS+CBA	20	50

1. SEN refers to special education needs.
2. England only.
3. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
4. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

[Part 2/2]
Table 12.3 **ETCS and TCS for countries/economies that participated in CBA and problem solving**

		CBA and PS	CBA Target Cluster Size (ETCS)	Target Cluster Size (TCS)
Partners	**Brazil**	PS+CBA	18 (only in the subsample of schools)	35
	Bulgaria	PS	14	35
	Colombia	PS+CBA	18	43
	Croatia	PS	14	43
	Cyprus[3, 4]	PS	35	80
	Hong Kong-China	PS+CBA	20	35
	Macao-China	PS+CBA	All students for all schools in strata 4, 8, and 9. 75 in strata 1, 2, 3, 5, 6, 7, and 10.	All
	Malaysia	PS	14	35
	Montenegro	PS	53 (14 in strata 1, 2, 3)	132 (25 in strata 1, 2, 3)
	Russian Federation	PS+CBA	18	43
	Serbia	PS	14	35
	Shanghai-China	PS+CBA	16	43
	Singapore	PS+CBA	18	35
	Chinese Taipei	PS+CBA	20	40
	United Arab Emirates	PS+CBA	18	35 in all strata except 61; 40 in strata 61
	Uruguay	PS	14	40

1. SEN refers to special education needs.
2. England only.
3. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
4. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

The international financial literacy cognitive data set (available as *FIN_COG12_MAR31.txt*[3]) consisted of 29 041 students from 18 participating countries and economies. Table 12.4 shows the total number of students included in the PISA 2012 financial literacy database, broken down by participating country/economy and test booklet.

Table 12.4 **Number of sampled students by country/economy and financial literacy booklet**

		Booklets					Total sampled students
		71	72	73	74	70	
OECD	**Australia**	819	871	864	739		3 293
	Belgium[1]	276	279	273	265		1 093
	Czech Republic	278	333	313	262	21	1 207
	Estonia	274	277	280	257		1 088
	France	252	296	289	231		1 068
	Israel	249	249	249	259		1 006
	Italy	1 772	1 793	1 784	1 719		7 068
	New Zealand	223	266	258	210		957
	Poland	250	270	273	261		1 054
	Slovak Republic	261	273	278	235	8	1 055
	Slovenia	286	363	343	292	28	1 312
	Spain	277	277	270	284		1 108
	United States	287	289	298	259		1 133
Partners	**Colombia**	515	528	542	515		2 100
	Croatia	282	286	283	294		1 145
	Latvia	241	246	250	233		970
	Russian Federation	292	311	306	278		1 187
	Shanghai-China	300	301	297	299		1 197

1. Flemish community only.

Test targeting

Calibration sample data for each of the domains was separately scaled to examine the targeting of the tests. Figures 12.1 to 12.6 show the match between the international item difficulty distribution and the distribution of student achievement for the paper-based assessments (PBA) of reading, mathematics, science and for the computer-based assessments (CBA) of mathematics, reading and problem solving, respectively. Figure 12.7 shows the distribution of student achievement for the assessment of financial literacy.

The figures consist of two panels. The first panel (students) shows the distribution of students' Rasch-scaled achievement estimates. Students at the top end of this distribution have higher proficiency estimates than the students at the lower end of the distribution. The second panel (item difficulties) shows the distribution of Rasch-estimated item difficulties.

A test is well targeted if the average of item difficulties is about the same as the average of the students' abilities and the item difficulties are evenly spread across the ability distribution.

In Figures 12.1 to 12.7, the distribution of student proficiency shown by *Xs*[4] is well matched to the distribution of item difficulty. The figures are constructed so that when a student and an item are at the same location on the scale, then the student has a 50% chance of responding correctly to the item.

■ Figure 12.1 ■

Item plot for paper-based mathematics items

```
Students         Item difficulties
-----------------------------------------
   4            |
                |106
                |
                |
                |
   3            |22 69
                |
               X|2 104
               X|
              XX|
              XX|88
   2          XX|3 14 57 87
              XX|72 84
             XXX|8 24 92 101 103
            XXXX|78
            XXXX|13 25 35 66 81
           XXXXX|39 40
   1     XXXXXXX|36 38 80
           XXXXX|47 67 83 94
          XXXXXX|15 20 59 74
          XXXXXX|1 5 10 48 75 109
        XXXXXXXX|33 44 54 107
   0    XXXXXXXX|16 17 31 32 42 76 91 99
      XXXXXXXXXX|18 30 56 68 77 96
        XXXXXXXX|9 11 27 64
       XXXXXXXXX|7 37 55 62 86 105
        XXXXXXXX|12 43 46 58 61 90 98
       XXXXXXXXX|6 21 60 63 89
  -1   XXXXXXXXX|23 28 29 65 73 79 95 108
      XXXXXXXXXX|49 82
        XXXXXXXX|4 26 85
        XXXXXXXX|41 52 100
           XXXXX|51 71 102
  -2       XXXXX|19
           XXXXX|53
            XXXX|
              XX|50 93
              XX|45
              XX|70 97
  -3           X|34
               X|
               X|
               X|
                |
                |
  -4            |
=========================================
Each ‘X’ represents 180.1 cases
=========================================
```

■ Figure 12.2 ■

Item plot for paper-based reading items

```
Students         Item difficulties
-----------------------------------------
   4            |
                |
                |
                |
                |
               X|
   3           X|
               X|25
              XX|43
               X|
              XX|28
              XX|
   2        XXXX|38
            XXXX|
           XXXXX|
         XXXXXXX|
           XXXXX|1 6
          XXXXXX|8 15 35
   1      XXXXXX|10
        XXXXXXXX|7 14 20 42
       XXXXXXXXX|17
      XXXXXXXXXX|5
        XXXXXXXX|26 27
        XXXXXXXX|3
   0    XXXXXXXX|2 13 33
        XXXXXXXX|32
        XXXXXXXX|19 21 37
        XXXXXXXX|9 24 34
         XXXXXXX|
         XXXXXXX|11 30
  -1        XXXX|4 22
            XXXX|18 31 36 44
            XXXX|40
             XXX|16 23 41
              XX|12
              XX|
  -2          XX|
               X|
               X|29
               X|
               X|
                |
  -3           X|
                |
                |39
                |
                |
                |
  -4            |
                |
                |
=========================================
Each ‘X’ represents 132.5 cases
=========================================
```

■ Figure 12.3 ■

Item plot for paper-based science items

```
Students         Item difficulties
-----------------------------------------
                |
                |
   3            |
                |
                |
               X|
                |
              XX|
   2           X|51
               X|
              XX|
             XXX|
            XXXX|10
             XXX|2 23 48
          XXXXXX|12
   1       XXXXX|30
          XXXXXX|6
          XXXXXX|17 33
        XXXXXXXX|15 41 46
          XXXXXX|5 14 22 27 31 37 44
        XXXXXXXX|1 40
      XXXXXXXXXX|21 53
   0     XXXXXXX|13 35 38 47 49 52
       XXXXXXXXX|4 9 20 32 45
         XXXXXXX|7 8 25 42
        XXXXXXXX|11 29
        XXXXXXXX|24
        XXXXXXXX|36 39
         XXXXXXX|16 19 26 34
  -1      XXXXXX|
            XXXX|18
            XXXX|28
           XXXXX|43
             XXX|
              XX|
  -2          XX|
              XX|50
               X|3
               X|
                |
               X|
                |
  -3            |
                |
                |
                |
                |
=========================================
Each 'X' represents 125.8 cases
=========================================
```

■ Figure 12.4 ■

Item plot for computer-based mathematics items

```
Students         Item difficulties
-----------------------------------------
                |
                |
   3            |1
                |
                |
                |
                |
                |
   2            |16
               X|37
               X|
              XX|38
              XX|14 17 35
              XX|
              XX|19 22
   1          XX|
            XXXX|41
            XXXX|20
           XXXXX|12 27 40
         XXXXXXX|9 13 25
           XXXXX|2 23
   0      XXXXXX|6 31
          XXXXXX|
        XXXXXXXX|3 18
        XXXXXXXX|4 11 28 33
      XXXXXXXXXX|5 7 29 36
        XXXXXXXX|26
        XXXXXXXX|
  -1    XXXXXXXX|8
       XXXXXXXXX|
        XXXXXXXX|21 30
       XXXXXXXXX|24 34
         XXXXXXX|10 15 39
         XXXXXXX|
  -2        XXXX|
           XXXXX|
            XXXX|32
             XXX|
              XX|
              XX|
  -3          XX|
               X|
               X|
               X|
               X|
                |
                |
  -4            |
                |
                |
                |
                |
=========================================
Each ‘X’ represents 88.2 cases
=========================================
```

■ Figure 12.5 ■

Item plot for digital reading items

```
Students         Item difficulties
-----------------------------------------
   4            |
                |
                |
                |
                |
               X|
   3           X|
               X|
              XX|
               X|
              XX|19
              XX|
            XXXX|
   2        XXXX|
           XXXXX|
         XXXXXXX|
           XXXXX|
          XXXXXX|
          XXXXXX|
   1    XXXXXXXX|12
        XXXXXXXX|
      XXXXXXXXXX|3 8 9
        XXXXXXXX|10 13 15
        XXXXXXXX|
         XXXXXXX|
        XXXXXXXX|16
   0    XXXXXXXX|11 18
        XXXXXXXX|5 7
          XXXXXX|6
          XXXXXX|4 17
            XXXX|
            XXXX|
            XXXX|
  -1         XXX|
              XX|2
              XX|
              XX|
               X|
               X|
  -2           X|
               X|14
                |
               X|
                |
                |1
                |
  -3            |
=========================================
Each ‘X’ represents 93.0 cases
=========================================
```

■ Figure 12.6 ■

Item plot for computer-based problem solving items

```
Students         Item difficulties
-----------------------------------------
                |
                |
   3            |1
                |
                |
                |
                |
                |
   2            |16
               X|37
               X|
              XX|38
              XX|14 17 35
              XX|
              XX|19 22
   1          XX|
            XXXX|41
            XXXX|20
           XXXXX|12 27 40
         XXXXXXX|9 13 25
           XXXXX|2 23
   0      XXXXXX|6 31
          XXXXXX|
        XXXXXXXX|3 18
        XXXXXXXX|4 11 28 33
      XXXXXXXXXX|5 7 29 36
        XXXXXXXX|26
        XXXXXXXX|
  -1    XXXXXXXX|8
       XXXXXXXXX|
        XXXXXXXX|21 30
       XXXXXXXXX|24 34
         XXXXXXX|10 15 39
         XXXXXXX|
  -2        XXXX|
           XXXXX|
            XXXX|32
             XXX|
              XX|
              XX|
  -3          XX|
               X|
               X|
               X|
               X|
                |
                |
  -4            |
                |
                |
                |
                |
                |
=========================================
Each ‘X’ represents 88.2 cases
=========================================
```

■ Figure 12.7 ■

Item plot for paper-based financial literacy items

```
Students          Item difficulties
-----------------------------------------
                |
   4            |
                |
                |
                |
                |
                |2
                |
                |
   3            |
                |32
               X|
               X|
               X|
              XX|
               X|
   2          XX|
              XX|
            XXXX|
            XXXX|
            XXXX|12 31
         XXXXXXX|6 27
           XXXXX|36
   1      XXXXXX|14 30 35
         XXXXXXX|22
        XXXXXXXX|37
       XXXXXXXXX|
      XXXXXXXXXX|8 17
        XXXXXXXX|5 19
       XXXXXXXXX|16 25 39
   0    XXXXXXXX|13
        XXXXXXXX|3 15
        XXXXXXXX|7 10
        XXXXXXXX|9 28
          XXXXXX|21 34
         XXXXXXX|26
             XXX|24
            XXXX|
  -1        XXXX|1 11 20 38
             XXX|
              XX|
              XX|
              XX|18
               X|23 29 33
               X|
  -2           X|
               X|40
                |
                |
                |
                |
                |4
  -3            |
                |
                |
                |
                |
                |
                |
=========================================
Each ‘X’ represents 17.0 cases
=========================================
```

Test reliability and measurement error design effect

A second test characteristic that is of importance is the test reliability, or equivalently the measurement error design effect (Adams, 2005). Table 12.5 shows the reliability of polled international equally weighted countries dataset for each of the three overall scales (mathematical literacy, reading literacy and scientific literacy), for the CBA scales (computer-based mathematics, digital reading and problem solving) and for financial literacy before conditioning and based upon seven separate unidimensional scalings, using plausible values (PV) and using Weighted Likelihood Estimates (WLE).

The WLE-based estimates are IRT analogues of traditional estimates of person separation reliability such as internal consistency. They are estimated for the samples of students that responded to test forms for each of the domains.

The plausible value-based estimates, however, use all students sampled for the paper-based test, and incorporate the influence of the test design on the uncertainty of estimates of the overall mean. For example, the digital reading reliability of 0.22 and corresponding design effect of 4.63 means that the error variance of the estimate of the mean would be increased by a factor of 4.63 because of the use of a sub-sample and 24 alternative assessment booklets. These estimates take into account the fact that the sample sizes for each domain are markedly different. The consequence is that the WLE reliabilities for the minor domains are higher than the PV reliabilities because students who were not assessed in reading, science or CBA domains were excluded from the calculation of the WLE reliabilities.

The plausible value-based estimates in Table 12.5 are based upon unidimensional scaling, and do not reflect the benefit of the conditioning and the multidimensional scaling that is implemented in PISA. The international reliability for each domain after conditioning and multidimensional scaling is reported in Table 12.11.

Table 12.5 **Reliabilities and Measurement Error Design Effect of each of the scales when scaled separately**

Domain	Reliability (WLE)	Measurement Error Design Effect (WLE)	Reliability (PV)	Measurement Error Design Effect (PV)
Mathematics	0.82	1.22	0.85	1.18
Reading	0.79	1.26	0.56	1.79
Science	0.80	1.25	0.57	1.76
Problem solving	0.79	1.27	0.26	3.92
Computer-based mathematics	0.73	1.37	0.20	4.95
Digital reading	0.74	1.35	0.22	4.63
Financial literacy	0.88	1.14	0.87	1.15

Domain inter-correlations

Correlations between the ability estimates for individual students in each of the three domains, the latent correlations, as estimated by *ConQuest* (Wu, Adams and Wilson, 1997) using calibration samples are given in Table 12.6. Correlations between six domains for countries that implemented CBA are given in Table 12.7. Correlations for three domains for the countries that implemented financial literacy are included in Table 12.8.

It is important to note that these latent correlations are unbiased estimates of the true correlation between the underlying latent variables. As such they are not attenuated by the unreliability of the measures and will generally be higher than the typical product moment correlations that have not been disattenuated for unreliability. The results in Table 12.6 are reported for both OECD countries and for all participating countries and economies, whereas the results in Tables 12.7 and 12.8 are reported only for all participating countries and economies.

Table 12.6 **Latent correlation between the three domains**

	Reading	Science
OECD countries	r	r
Mathematics	0.85	0.89
Reading		0.85

	Reading	Science
All countries/economies	r	r
Mathematics	0.86	0.90
Reading		0.88

Table 12.7 **Latent correlation between the six domains**

All countries/economies	Print reading r	Print science r	Problem solving r	Computer-based mathematics r	Digital reading r
Print mathematics	0.89	0.91	0.85	0.93	0.82
Print reading		0.91	0.82	0.83	0.86
Print science			0.80	0.88	0.82
Problem solving				0.86	0.81
Computer-based mathematics					0.91

Table 12.8 **Latent correlation between the three domains (financial literacy)**

All countries/economies	Reading r	Financial literacy r
Mathematics	0.84	0.82
Reading		0.76

Mathematics scale and subscales

As described in Chapter 9, a six-dimensional model consisting of reading, science, and the four mathematics content subscales *change and relationships*, *quantity*, *space and shape*, *uncertainty and data* was used. Then a five-dimensional model was estimated consisting of reading, science, and the three mathematics process subscales: *formulate*, *employ* and *interpret*. Responses from the reading and science domains were included in the scaling model to improve the estimation of posterior distributions of the mathematics subscales. The plausible values for reading and science generated using these two models were not included in the international database. The correlations between mathematics subscales as estimated from these two models are given in Table 12.9.

Table 12.9 **Latent correlation between mathematics subscales**

All countries/economies	Interpreting	Formulating	Quantity	Uncertainty and data	Space and shape
Formulating	0.96				
Employing	0.96	0.97			
Uncertainty and data			0.87		
Space and shape			0.84	0.85	
Change and relationships			0.91	0.92	0.90

SCALING OUTCOMES

The procedures for the national and international scaling are outlined in Chapter 9 and are not reiterated here.

National item deletions

The items were first scaled by country and their fit was considered at the national level, as was the consistency of the item parameter estimates across countries. Consortium staff then adjudicated items, considering the items' functioning both within and across countries in detail. Those items considered to be dodgy (see Chapter 9) were then reviewed in consultation with National Project Managers (NPMs). The consultations resulted in the deletion of a number of items at the national level.

At the international level, one mathematics item (*PM603Q02*) was deleted from scaling. *PM603Q02* was deleted because of concerns that arose during data adjudication regarding the degree of consistency with which the intended coding rules were applied across countries. The nationally deleted items are listed in Table 12.10. All deleted items were recoded as not applicable and were excluded from both international scaling and generation of plausible values.

Table 12.10 Items deleted at the national level

Item	Country/Economy
PM411Q02	United Arab Emirates (Arabic-language version)
PM828Q01	Viet Nam
PM909Q01	Jordan (Arabic-language version)
PM949Q03	Slovak Republic (booklet 20)
PM954Q02	Montenegro (Montenegrin-language versions)
PM985Q03	Brazil
PR404Q07T	Greece, Cyprus[1, 2] (Greek-language version)
PR406Q01	Finland (Finnish-language version)
PR420Q02	Spain (Catalan-language version)
PR420Q10	Finland
PR437Q01	Viet Nam
PR437Q06	Iceland
PR453Q06	Iceland
PR455Q05	Austria (German-language version), Belgium (German-language version), Switzerland (German-language version), Germany, Italy (German-language version), Luxembourg (German-language version), Liechtenstein
PS131Q02	France (booklet 1)
PS131Q04	France (booklet 1)
PS131Q04D	Mexico (booklet 10)
PS326Q02	Brazil
PS415Q08	Austria (German-language version), Belgium (German-language version), Switzerland (German-language version), Germany, Italy (German-language version), Luxembourg (German-language version), Liechtenstein
PS498Q03	Qatar (Arabic-language version)
PS519Q02	Montenegro (Montenegrin-language versions)
CM005Q02	Colombia
CM006Q03	Austria (German-language version)
CM011Q01T	United Arab Emirates (Arabic-language version), Israel (Hebrew and Arabic-language versions)
CM014Q02	Colombia
CM020Q01	Chile
CM020Q03	Chile
CP018Q05	France
CP034Q02	United Arab Emirates (Arabic-language version)
CR014Q06	United Arab Emirates (Arabic-language version), Brazil
CR014Q11	United Arab Emirates (Arabic-language version)
CR017Q01	Norway
CR017Q04	Israel
CR017Q07	Hong Kong-China (Chinese-language version), Macao-China (Chinese-language version), Shanghai-China
CR021Q08	United Arab Emirates (Arabic-language version), Colombia, Spain (Catalan-language version), Estonia (Russian-language version), Hungary, Israel (Hebrew and Arabic-language versions), Portugal
PF004Q03	Shanghai-China
PF055Q03	Shanghai-China
PF106Q02	Slovak Republic (Booklet 71 Slovenian -language version)

1. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
2. Footnote by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

International scaling

The international scaling for mathematics, science and reading items were performed using a calibration data set of 31 500 students (500 randomly selected students from each of the participating countries and economies, except for Cyprus[5] and Liechtenstein).

The item parameter estimates from this scaling are reported in Annex A. The item parameters were estimated using three separate one-dimensional models. As in previous cycles, not-reached items were treated as not administered and a booklet facet was used in the item response model.

The international scaling for CBA items was performed using a calibration data set of 29 568 students. The process of selecting the calibration sample for each of the PISA 2012 CBA domains is described in Chapter 9. The item parameter estimates from this scaling are reported in Annex A.

The international scaling for financial literacy items was performed using a calibration data set of 2 772 students (154 randomly selected students from each of the 18 participating countries and economies). The item parameter estimates from this scaling are reported in Annex A.

Generating student scale scores and reliability of the PISA scales

Applying the conditioning approach described in Chapter 9 and anchoring all of the item parameters at the values obtained from the international scaling, plausible values were generated for all sampled students. Table 12.11 gives the average of national reliabilities for the generated scale scores. The increase in reliability of the results reported in Table 12.11 over those presented in Table 12.5 is due to the use of multidimensional scaling and conditioning.

Table 12.11 **Final reliabilities of the PISA scales and subscales**

Domains	Reliability
Mathematics	0.914
Reading	0.888
Science	0.885
Problem solving	0.878
Financial literacy	0.937
Mathematics subscales	
Change and relationships	0.911
Quantity	0.903
Space and shape	0.891
Uncertainty and data	0.904
Employing	0.909
Formulating	0.892
Interpreting	0.897
Computer-based assessments	
Computer-based mathematics	0.925
Digital reading	0.913

Table 12.12 gives the reliabilities at the national level for the generated scale scores.

[Part 1/2]

Table 12.12 National reliabilities of the main domains and mathematics subscales

		Mathematics	Reading	Science	Problem solving	Financial literacy	Mathematics subscales	
							Change and relationships	Quantity
OECD	Australia	0.94	0.92	0.92	0.88	0.95	0.93	0.92
	Austria	0.93	0.92	0.92	0.89		0.93	0.91
	Belgium	0.94	0.93	0.93	0.93	0.94[1]	0.95	0.93
	Canada	0.92	0.89	0.90	0.92		0.91	0.90
	Chile	0.91	0.87	0.87	0.90		0.90	0.89
	Czech Republic	0.94	0.90	0.89	0.90	0.95	0.94	0.92
	Denmark	0.92	0.90	0.91	0.90		0.91	0.90
	Estonia	0.91	0.90	0.88	0.88	0.92	0.91	0.89
	Finland	0.92	0.88	0.88	0.83		0.91	0.90
	France	0.94	0.91	0.92	0.91	0.95	0.93	0.92
	Germany	0.94	0.93	0.92	0.93		0.93	0.91
	Greece	0.90	0.86	0.86			0.90	0.89
	Hungary	0.94	0.92	0.92	0.91		0.94	0.93
	Iceland	0.91	0.88	0.88			0.91	0.91
	Ireland	0.92	0.90	0.91	0.88		0.93	0.92
	Israel	0.93	0.90	0.91	0.92	0.94	0.92	0.92
	Italy	0.93	0.91	0.91	0.95	0.95	0.93	0.93
	Japan	0.93	0.89	0.89	0.77		0.91	0.91
	Korea	0.93	0.90	0.90	0.89		0.91	0.91
	Luxembourg	0.92	0.89	0.89			0.91	0.90
	Mexico	0.88	0.85	0.84			0.88	0.88
	Netherlands	0.94	0.92	0.91	0.89		0.95	0.93
	New Zealand	0.93	0.91	0.91		0.95	0.93	0.93
	Norway	0.92	0.89	0.90	0.86		0.92	0.90
	Poland	0.93	0.89	0.88	0.89	0.92	0.93	0.91
	Portugal	0.93	0.90	0.90	0.91		0.93	0.92
	Slovak Republic	0.94	0.93	0.93	0.92	0.95	0.93	0.93
	Slovenia	0.93	0.92	0.91	0.88	0.95	0.93	0.92
	Spain	0.92	0.89	0.90	0.90	0.93	0.93	0.92
	Sweden	0.92	0.89	0.90	0.85		0.92	0.90
	Switzerland	0.92	0.90	0.90			0.92	0.92
	Turkey	0.92	0.89	0.87	0.86		0.92	0.91
	United Kingdom	0.93	0.91	0.92	0.87[2]		0.93	0.92
	United States	0.93	0.91	0.92	0.89	0.94	0.93	0.92
	United States (Connecticut)	0.93	0.90	0.91			0.92	0.93
	United States (Florida)	0.92	0.87	0.89			0.93	0.92
	United States (Massachusetts)	0.93	0.89	0.90			0.94	0.93
Partners	Albania	0.86	0.81	0.83			0.84	0.85
	Argentina	0.89	0.87	0.87			0.89	0.89
	Brazil	0.89	0.86	0.85	0.88		0.90	0.89
	Bulgaria	0.92	0.91	0.90	0.85		0.89	0.90
	Colombia	0.88	0.88	0.85	0.92	0.93	0.87	0.88
	CostaRica	0.87	0.83	0.81			0.90	0.89
	Croatia	0.92	0.91	0.88	0.87	0.94	0.93	0.91
	Cyprus[3, 4]	0.91	0.88	0.88	0.82		0.91	0.90
	Hong Kong-China	0.92	0.91	0.90	0.84		0.92	0.91
	Indonesia	0.85	0.83	0.81			0.87	0.85
	Jordan	0.88	0.88	0.85			0.87	0.86
	Kazakhstan	0.84	0.84	0.80			0.81	0.85
	Latvia	0.91	0.89	0.88		0.93	0.91	0.91
	Liechtenstein	0.92	0.90	0.89			0.92	0.92
	Lithuania	0.91	0.90	0.88			0.92	0.91
	Macao-China	0.92	0.87	0.87	0.78		0.90	0.89
	Malaysia	0.90	0.87	0.87	0.85		0.90	0.90
	Montenegro	0.89	0.87	0.86	0.79		0.88	0.87
	Peru	0.89	0.86	0.84			0.90	0.89
	Qatar	0.90	0.87	0.87			0.90	0.89
	Romania	0.89	0.86	0.86			0.88	0.87
	Russian Federation	0.91	0.88	0.87	0.87	0.91	0.90	0.90
	Russian Federation (Perm)	0.91	0.89	0.89			0.93	0.92
	Serbia	0.92	0.88	0.87	0.85		0.90	0.89
	Shanghai-China	0.92	0.90	0.89	0.89	0.93	0.90	0.89
	Singapore	0.94	0.91	0.93	0.88		0.93	0.92
	Chinese Taipei	0.94	0.91	0.91	0.89		0.93	0.92
	Thailand	0.89	0.87	0.84			0.89	0.89
	Tunisia	0.88	0.84	0.81			0.86	0.87
	United Arab Emirates	0.92	0.91	0.91	0.89		0.91	0.91
	Uruguay	0.91	0.88	0.86	0.85		0.89	0.89
	Viet Nam	0.91	0.86	0.87			0.90	0.90

1. Flemish community only.
2. England only.
3. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
4. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

[Part 2/2]

Table 12.12 National reliabilities of the main domains and mathematics subscales

		Mathematics subscales					Computer-based assessments	
		Space and Shape	Uncertainty and Data	Employing	Formulating	Interpreting	Computer-based mathematics	Digital reading
OECD	Australia	0.92	0.93	0.93	0.92	0.93	0.93	0.91
	Austria	0.90	0.92	0.93	0.91	0.91	0.94	0.93
	Belgium	0.92	0.94	0.94	0.92	0.93	0.96	0.95
	Canada	0.90	0.91	0.91	0.91	0.91	0.94	0.95
	Chile	0.87	0.88	0.90	0.87	0.88	0.92	0.92
	Czech Republic	0.92	0.93	0.93	0.93	0.93		
	Denmark	0.91	0.90	0.90	0.90	0.90	0.92	0.91
	Estonia	0.88	0.89	0.90	0.89	0.88	0.92	0.90
	Finland	0.89	0.89	0.91	0.90	0.90		
	France	0.92	0.93	0.93	0.92	0.92	0.95	0.94
	Germany	0.92	0.93	0.94	0.92	0.93	0.95	0.95
	Greece	0.87	0.90	0.89	0.88	0.89		
	Hungary	0.92	0.93	0.93	0.91	0.91	0.94	0.94
	Iceland	0.92	0.90	0.91	0.90	0.90		
	Ireland	0.91	0.91	0.91	0.90	0.89	0.93	0.92
	Israel	0.92	0.92	0.92	0.92	0.92	0.94	0.94
	Italy	0.92	0.93	0.93	0.92	0.92	0.96	0.96
	Japan	0.89	0.91	0.92	0.91	0.90	0.84	0.82
	Korea	0.92	0.92	0.93	0.92	0.91	0.92	0.90
	Luxembourg	0.89	0.91	0.92	0.89	0.90		
	Mexico	0.87	0.88	0.88	0.86	0.88		
	Netherlands	0.93	0.93	0.94	0.93	0.91		
	New Zealand	0.91	0.92	0.92	0.91	0.91		
	Norway	0.91	0.91	0.92	0.91	0.90	0.90	0.89
	Poland	0.90	0.92	0.92	0.91	0.90	0.94	0.90
	Portugal	0.91	0.93	0.92	0.91	0.90	0.95	0.95
	Slovak Republic	0.92	0.93	0.93	0.92	0.92	0.94	0.93
	Slovenia	0.90	0.93	0.94	0.92	0.93	0.94	0.91
	Spain	0.92	0.92	0.92	0.91	0.90	0.94	0.94
	Sweden	0.90	0.91	0.90	0.90	0.90	0.91	0.91
	Switzerland	0.89	0.92	0.92	0.91	0.91		
	Turkey	0.88	0.92	0.91	0.90	0.90		
	United Kingdom	0.91	0.92	0.92	0.91	0.91		
	United States	0.88	0.92	0.91	0.90	0.91	0.92	0.92
	United States (Connecticut)	0.91	0.93	0.93	0.91	0.91		
	United States (Florida)	0.89	0.91	0.92	0.89	0.90		
	United States (Massachusetts)	0.93	0.93	0.93	0.93	0.90		
Partners	Albania	0.81	0.86	0.85	0.81	0.88		
	Argentina	0.87	0.87	0.89	0.86	0.89		
	Brazil	0.88	0.89	0.89	0.88	0.89	0.91	0.92
	Bulgaria	0.90	0.90	0.90	0.88	0.90		
	Colombia	0.86	0.87	0.88	0.86	0.89	0.92	0.90
	CostaRica	0.89	0.88	0.88	0.85	0.85		
	Croatia	0.91	0.92	0.92	0.90	0.90		
	Cyprus[1, 2]	0.88	0.90	0.92	0.88	0.90		
	Hong Kong-China	0.91	0.90	0.92	0.90	0.90	0.92	0.89
	Indonesia	0.82	0.86	0.86	0.82	0.85		
	Jordan	0.85	0.87	0.88	0.85	0.86		
	Kazakhstan	0.81	0.86	0.83	0.81	0.83		
	Latvia	0.91	0.91	0.91	0.91	0.90		
	Liechtenstein	0.88	0.91	0.92	0.91	0.92		
	Lithuania	0.90	0.91	0.92	0.90	0.90		
	Macao-China	0.88	0.87	0.91	0.89	0.87	0.84	0.81
	Malaysia	0.87	0.88	0.91	0.87	0.88		
	Montenegro	0.84	0.87	0.88	0.85	0.86		
	Peru	0.88	0.90	0.90	0.85	0.90		
	Qatar	0.85	0.89	0.89	0.86	0.89		
	Romania	0.84	0.86	0.89	0.84	0.86		
	Russian Federation	0.88	0.91	0.90	0.88	0.89	0.90	0.88
	Russian Federation (Perm)	0.89	0.91	0.90	0.90	0.90		
	Serbia	0.88	0.90	0.91	0.89	0.89		
	Shanghai-China	0.89	0.90	0.91	0.90	0.88	0.92	0.90
	Singapore	0.89	0.90	0.93	0.91	0.91	0.93	0.92
	Chinese Taipei	0.92	0.92	0.93	0.93	0.90	0.91	0.87
	Thailand	0.88	0.89	0.89	0.87	0.88		
	Tunisia	0.84	0.88	0.88	0.84	0.86		
	United Arab Emirates	0.88	0.91	0.91	0.89	0.91	0.95	0.94
	Uruguay	0.88	0.88	0.90	0.88	0.89		
	Viet Nam	0.88	0.89	0.91	0.89	0.89		

1. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

2. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Test length analysis

Numbers of missing responses are discussed in this section, with a differentiation between item level non-responses and not-reached responses. A response is coded as item level non-response if the student was expected to answer a question, but no response was actually provided. All consecutive missing values clustered at the end of a test session were replaced by the not-reached code, except for the first value of the missing series, which is coded as item level non-response (see Chapter 19). All the tables included in the section include weighted and unweighted numbers of the item level non-responses and not-reached responses. The final student weight (see Chapter 8) was used to provide weighted numbers and percents.

Table 12.13 shows the number of missing responses and the number of missing responses recoded as not reached, by booklet. Tables 12.14 and 12.15 show the number of missing and not reached responses by CBA test form and financial literacy booklet, respectively.

Table 12.13 **Average number of not-reached items and missing items by booklet**

	Missing		Not reached	
Booklet	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
1	5.53	4.47	1.12	1.09
2	4.90	4.01	1.22	1.20
3	4.96	4.29	0.71	0.73
4	5.23	4.37	0.77	0.71
5	5.07	4.26	0.86	0.79
6	5.45	4.66	0.78	0.63
7	5.02	4.10	1.08	0.97
8	4.46	3.93	1.75	1.88
9	4.87	4.30	1.33	1.34
10	5.62	5.06	1.01	1.11
11	4.14	3.74	0.78	0.76
12	5.03	4.56	1.89	1.95
13	5.05	4.73	1.99	2.17
21	5.63	5.49	3.44	3.69
22	5.22	5.10	3.53	3.98
23	5.56	5.54	2.25	2.45
24	4.70	4.53	1.93	2.24
25	5.47	5.48	2.26	2.55
26	5.98	5.98	2.51	2.69
27	5.64	5.58	2.42	2.58
UH	3.67	4.01	0.63	0.77
Total	**5.07**	**4.51**	**1.44**	**1.49**

Average number of missing and not-reached items could be compared between standard booklets 1 to 7 and non-standard booklets 21 to 27. On average, standard booklets have fewer not-reached items and less missing data.

Average numbers of missing and not-reached items for CBA are significantly lower compared to the PBA.

Table 12.14 **Average number of not-reached items and missing items by computer-based test form**

	Missing		Not reached	
Form ID	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
31	1.23	0.75	0.20	0.09
32	1.37	0.88	0.10	0.05
33	1.16	0.79	0.05	0.02
34	0.76	0.51	0.09	0.06
35	1.21	0.82	0.12	0.04
36	1.51	1.12	0.23	0.12
37	0.93	0.64	0.08	0.03
38	0.92	0.56	0.05	0.03
41	0.78	0.72	0.10	0.12
42	0.61	0.64	0.04	0.03
43	1.65	1.70	0.17	0.15
44	0.90	0.87	0.07	0.05
45	1.32	1.32	0.14	0.12
46	0.92	0.91	0.03	0.03
47	0.61	0.63	0.04	0.03
48	1.62	1.66	0.19	0.20
49	0.77	0.69	0.03	0.01
50	1.05	1.06	0.07	0.04
51	1.29	1.26	0.21	0.16
52	0.80	0.73	0.09	0.06
53	1.62	1.63	0.24	0.15
54	0.51	0.48	0.07	0.04
55	0.59	0.68	0.03	0.03
56	1.07	1.03	0.19	0.16
57	0.86	0.87	0.05	0.03
58	1.24	1.21	0.06	0.04
59	0.89	0.85	0.07	0.03
60	1.65	1.64	0.21	0.16
61	0.62	0.64	0.04	0.03
62	1.46	1.35	0.08	0.07
63	1.02	1.02	0.19	0.18
64	0.53	0.50	0.03	0.02
Total	**0.58**	**0.57**	**0.06**	**0.04**

Table 12.15 **Average number of not-reached items and missing items by financial literacy booklet**

	Missing		Not reached	
Booklet	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
1	4.99	3.73	2.05	2.32
2	5.09	3.92	1.96	1.82
3	5.00	3.75	1.86	2.02
4	4.95	3.77	2.54	2.86
5	3.53	3.17	0.16	0.02
Total	**5.01**	**3.79**	**2.09**	**2.24**

Table 12.16 shows the number of not-reached items for the paper–based assessment by country. Tables 12.17 and 12.18 show this information by country over all CBA test forms and financial literacy booklets, respectively. The average number of not-reached items differs from one country to another. Generally, countries with higher averages of not-reached items also have higher averages of missing data.

Table 12.16 **Average number of not-reached items and missing items by country/economy for PBA in the main domains**

		Missing		Not reached	
		Unweighted	Weighted	Unweighted	Weighted
OECD	**Australia**	3.53	4.03	0.57	0.73
	Austria	5.67	5.73	0.35	0.44
	Belgium	4.11	4.04	0.80	0.74
	Canada	2.79	3.09	0.97	0.99
	Chile	6.32	5.46	1.71	1.54
	Czech Republic	5.37	4.67	0.48	0.39
	Denmark	4.30	4.92	0.69	0.84
	Estonia	3.44	3.43	0.46	0.44
	Finland	3.64	4.23	0.56	0.77
	France	6.16	5.99	1.39	1.34
	Germany	4.87	4.87	0.37	0.38
	Greece	6.85	6.82	1.55	1.53
	Hungary	5.30	5.06	0.60	0.51
	Iceland	4.92	4.93	1.19	1.17
	Ireland	3.12	3.08	0.39	0.38
	Israel	6.41	6.22	2.27	2.10
	Italy	6.71	6.29	1.04	0.99
	Japan	4.25	4.28	0.44	0.45
	Korea	2.52	2.56	0.26	0.26
	Luxembourg	5.29	5.26	0.96	0.96
	Mexico	3.48	3.43	3.71	3.27
	Netherlands	1.60	1.65	0.18	0.17
	New Zealand	4.30	4.28	0.84	0.83
	Norway	5.99	6.02	1.27	1.21
	Poland	3.88	3.77	0.31	0.30
	Portugal	5.23	5.38	1.06	1.13
	Slovak Republic	6.52	6.30	0.57	0.54
	Slovenia	5.41	6.34	0.37	0.49
	Spain	5.59	5.10	1.10	0.85
	Sweden	6.29	6.21	1.77	1.76
	Switzerland	4.10	4.42	0.50	0.53
	Turkey	3.27	3.26	0.46	0.44
	United Kingdom	4.31	4.39	0.65	0.52
	United States	1.68	1.62	0.56	0.56
	United States (Connecticut)	1.83	1.88	0.35	0.36
	United States (Florida)	1.77	1.82	0.45	0.46
	United States (Massachusetts)	1.48	1.63	0.54	0.63
Partners	**Albania**	13.19	13.23	3.54	3.34
	Argentina	10.07	9.99	2.68	2.92
	Brazil	6.16	6.42	3.31	3.55
	Bulgaria	8.41	8.18	1.11	1.07
	Colombia	6.16	5.93	5.57	5.04
	CostaRica	4.07	4.06	2.78	2.76
	Croatia	6.53	6.55	0.44	0.43
	Cyprus[1, 2]	7.69	8.13	1.43	1.48
	Hong Kong-China	2.30	2.31	0.39	0.40
	Indonesia	5.39	5.28	1.53	1.46
	Jordan	6.16	6.30	1.82	1.74
	Kazakhstan	5.45	5.42	1.04	1.04
	Latvia	3.58	3.53	0.48	0.51
	Liechtenstein	3.65	3.55	0.34	0.33
	Lithuania	4.72	4.70	0.39	0.39
	Macao-China	2.55	2.55	0.86	0.86
	Malaysia	6.15	6.11	1.30	1.27
	Montenegro	11.27	11.37	1.08	1.05
	Peru	7.10	7.11	5.84	5.80
	Qatar	8.94	8.94	3.20	3.20
	Romania	1.02	0.98	0.03	0.02
	Russian Federation	6.39	6.44	1.52	1.51
	Russian Federation (Perm)	6.14	6.04	1.23	1.17
	Serbia	9.09	9.19	1.23	1.22
	Shanghai-China	1.31	1.31	0.10	0.10
	Singapore	2.09	2.20	0.46	0.52
	Chinese Taipei	3.14	3.17	0.25	0.27
	Thailand	4.03	3.91	1.10	1.12
	Tunisia	8.21	8.16	3.77	3.79
	United Arab Emirates	3.80	3.90	1.37	1.31
	Uruguay	9.47	9.39	4.70	4.59
	Viet Nam	2.06	2.05	0.06	0.04

Note: Countries which implemented the easier set of booklets are shown in italics.

1. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

2. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Table 12.17 Average number of not-reached items and missing items by country/economy for CBA

		Missing		Not reached	
		Unweighted	Weighted	Unweighted	Weighted
OECD	Australia	0.73	0.67	0.09	0.07
	Austria	0.76	0.75	0.07	0.07
	Belgium	0.59	0.66	0.06	0.07
	Canada	0.33	0.32	0.02	0.02
	Chile	0.72	0.83	0.06	0.06
	Czech Republic	0.39	0.56	0.01	0.02
	Denmark	0.62	0.53	0.07	0.05
	Estonia	0.46	0.51	0.03	0.03
	Finland	0.28	0.26	0.01	0.01
	France	0.68	0.70	0.06	0.06
	Germany	0.71	0.71	0.07	0.06
	Hungary	0.71	0.81	0.07	0.08
	Ireland	0.45	0.45	0.02	0.03
	Israel	0.84	0.90	0.22	0.27
	Italy	0.58	0.55	0.05	0.04
	Japan	1.04	1.04	0.06	0.06
	Korea	0.32	0.32	0.02	0.02
	Netherlands	0.29	0.27	0.02	0.02
	Norway	0.69	0.70	0.11	0.12
	Poland	0.65	0.66	0.06	0.06
	Portugal	0.47	0.47	0.04	0.04
	Slovak Republic	0.72	0.77	0.05	0.05
	Slovenia	1.13	0.93	0.14	0.09
	Spain	0.70	0.76	0.08	0.11
	Sweden	0.72	0.73	0.07	0.07
	Turkey	0.29	0.29	0.01	0.01
	United Kingdom[1]	0.25	0.26	0.03	0.03
	United States	0.28	0.29	0.02	0.02
Partners	Brazil	0.76	0.76	0.06	0.06
	Bulgaria	0.95	1.06	0.14	0.16
	Colombia	0.74	0.77	0.08	0.07
	Croatia	0.42	0.42	0.04	0.04
	Cyprus[2, 3]	0.79	0.69	0.09	0.07
	Hong Kong-China	0.32	0.32	0.03	0.02
	Macao-China	0.30	0.30	0.01	0.01
	Malaysia	0.34	0.34	0.02	0.02
	Montenegro	0.81	0.78	0.08	0.09
	Russian Federation	0.74	0.79	0.05	0.06
	Serbia	0.41	0.43	0.03	0.03
	Shanghai-China	0.20	0.20	0.01	0.01
	Singapore	0.22	0.21	0.02	0.02
	Chinese Taipei	0.34	0.33	0.02	0.02
	United Arab Emirates	0.78	0.73	0.11	0.11
	Uruguay	0.98	1.03	0.16	0.17

1. England only.
2. Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
3. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Table 12.18 Average number of not-reached items and missing items by country/economy for the financial literacy assessment

		Missing		Not reached	
		Unweighted	Weighted	Unweighted	Weighted
OECD	Australia	3.48	2.97	1.10	0.81
	Belgium[1]	2.27	2.38	0.81	0.84
	Czech Republic	4.60	5.35	0.89	0.97
	Estonia	2.79	2.78	0.74	0.74
	France	5.34	5.51	2.33	2.43
	Israel	5.49	5.69	3.67	4.14
	Italy	6.91	7.29	2.07	2.10
	New Zealand	4.00	4.01	1.65	1.74
	Poland	3.63	3.75	0.65	0.66
	Slovak Republic	6.99	7.14	1.06	1.03
	Slovenia	6.57	5.43	0.81	0.69
	Spain	5.52	5.68	2.13	2.08
	United States	1.47	1.46	0.99	1.10
Partners	Colombia	5.77	6.02	9.14	9.84
	Croatia	6.26	6.23	0.88	0.86
	Latvia	3.29	3.26	1.28	1.24
	Russian Federation	6.50	6.53	3.35	3.13
	Shanghai-China	1.00	0.98	0.13	0.12

1. Flemish community only.

Tables 12.19, 12.20 and 12.21 provide the percentage distribution of not-reached items per booklet and CBA test form. The percentage of students who reached the last item (i.e. the percentages of students with zero not-reached items) for PBA ranges from 71% to 94% when using weighted data and 72% to 94% when using unweighted data. The percentage of students who reached the last item for the CBA ranges from 89% to nearly 100% when using weighted data and from 87% to 99% when using unweighted data. More than 80% of students reached the last item for financial literacy assessment.

Table 12.19 Distribution of not-reached items by booklet

	Number of non-reached items									
	0	1	2	3	4	5	6	7	8	>8
Booklet	Weighted percentages									
1	88.39	0.85	1.25	1.96	0.62	0.61	0.63	0.88	0.24	4.56
2	87.73	1.14	2.06	0.26	0.89	0.65	0.82	0.86	0.12	5.47
3	93.88	0.36	0.18	0.39	0.65	0.19	0.30	0.53	0.24	3.27
4	86.70	6.34	1.49	0.70	0.24	0.42	0.28	0.81	0.31	2.69
5	91.08	0.96	0.62	0.49	0.68	0.44	1.61	0.15	0.39	3.59
6	91.00	3.34	0.14	0.45	0.57	0.24	0.34	1.19	0.13	2.60
7	82.36	3.61	2.43	4.94	0.66	0.40	0.66	1.04	0.40	3.51
8	86.11	0.58	0.78	0.50	0.37	0.92	0.95	0.37	1.03	8.37
9	88.35	0.61	0.37	0.75	0.67	0.46	1.08	0.16	1.32	6.23
10	85.03	2.00	2.56	2.21	0.82	1.49	0.26	0.43	0.20	4.99
11	90.32	0.37	1.69	1.64	0.49	1.04	0.64	0.27	0.42	3.13
12	83.74	0.88	0.61	0.69	1.63	0.59	0.76	0.61	0.55	9.94
13	79.59	0.90	3.73	1.75	0.78	1.42	0.47	0.48	1.43	9.45
21	73.01	1.11	1.80	2.99	1.05	1.13	0.71	1.01	0.63	16.56
22	71.44	1.04	2.59	0.43	1.62	0.62	1.36	1.45	0.74	18.71
23	81.66	0.61	0.43	0.40	1.92	0.67	0.75	1.59	0.77	11.20
24	73.15	8.89	1.56	1.55	0.93	1.37	0.68	1.24	0.71	9.93
25	76.67	1.46	0.89	0.77	1.87	1.19	2.65	0.35	1.06	13.11
26	76.84	2.00	2.11	0.95	0.78	0.98	1.94	1.08	1.57	11.76
27	75.68	4.20	0.22	1.19	1.18	1.15	1.09	1.21	2.05	12.04
UH	88.34	2.35	0.70	0.93	0.62	1.63	0.36	0.37	0.24	4.46
	Unweighted percentages									
1	88.04	1.02	1.49	2.12	0.70	0.56	0.57	0.62	0.21	4.69
2	87.65	1.17	2.45	0.23	1.18	0.49	0.70	0.68	0.19	5.27
3	93.89	0.34	0.43	0.39	0.82	0.19	0.26	0.67	0.24	2.77
4	84.93	8.13	1.29	0.73	0.28	0.63	0.42	0.65	0.21	2.72
5	90.01	1.11	0.90	0.58	0.82	0.55	1.63	0.18	0.43	3.78
6	89.24	4.13	0.17	0.53	0.49	0.34	0.47	1.21	0.24	3.17
7	78.15	4.76	2.89	7.02	0.77	0.49	0.67	1.29	0.57	3.38
8	86.71	0.58	0.98	0.50	0.45	0.92	0.85	0.37	1.10	7.55
9	88.46	0.54	0.42	0.86	0.70	0.51	1.08	0.17	1.46	5.81
10	85.70	2.43	2.53	2.05	0.85	1.31	0.21	0.38	0.20	4.33
11	89.78	0.47	1.89	1.94	0.56	1.09	0.65	0.26	0.41	2.95
12	84.11	1.05	0.63	0.74	1.67	0.51	0.81	0.77	0.67	9.04
13	79.73	0.95	4.43	1.75	0.87	1.38	0.36	0.58	1.23	8.72
21	72.96	1.19	1.90	3.32	1.41	1.13	0.76	1.37	0.57	15.39
22	72.31	1.31	3.16	0.40	1.88	0.78	1.21	1.35	0.68	16.93
23	82.09	0.73	0.59	0.51	1.99	0.57	0.61	1.82	0.82	10.26
24	74.67	9.01	1.72	1.37	0.81	1.41	0.81	1.34	0.57	8.30
25	77.39	1.39	1.06	1.04	1.98	1.33	3.14	0.46	0.95	11.27
26	77.29	2.10	2.21	1.05	0.76	1.01	2.18	1.03	1.65	10.71
27	75.77	3.97	0.23	1.31	1.04	1.36	1.30	1.08	2.42	11.51
UH	89.52	2.13	0.89	0.89	0.80	1.95	0.62	0.18	0.62	2.40

Table 12.20 Distribution of not-reached items by computer-based test form

	Number of non-reached items									
	0	1	2	3	4	5	6	7	8	>8
Booklet	Weighted percentages									
31	96.37	2.06	0.85	0.12	0.02	0.14	0.05	0.09	0.02	0.28
32	98.19	1.17	0.13	0.16	0.03	0.01	0.04	0.09	0.01	0.17
33	99.48	0.12	0.03	0.17		0.14		0.03		0.04
34	97.90	1.41	0.24	0.01	0.07	0.01	0.03		0.10	0.22
35	97.94	1.53	0.31	0.04	0.06		0.04	0.01	0.03	0.05
36	94.70	2.68	1.07	0.84	0.20	0.08	0.13	0.01	0.02	0.26
37	98.27	1.30	0.17	0.05	0.03	0.07	0.04	0.01	0.02	0.05
38	99.25	0.30	0.11	0.12		0.13		0.01	0.01	0.08
41	97.08	1.67	0.26	0.18	0.00	0.03	0.24		0.00	0.54
42	98.45	1.16	0.20	0.06	0.06		0.01	0.00	0.00	0.06
43	92.23	4.48	1.90	0.52	0.24	0.18	0.28	0.08	0.01	0.06
44	98.73	0.66	0.03	0.05	0.02	0.06	0.12	0.17	0.03	0.11
45	89.93	9.51	0.27	0.03	0.02	0.04	0.03	0.00	0.12	0.03
46	99.40	0.17	0.03	0.01	0.30	0.05				0.04
47	99.12	0.46	0.22	0.04	0.04	0.00	0.01	0.03	0.01	0.07
48	88.81	7.60	1.57	1.05	0.08	0.76	0.00	0.04	0.02	0.07
49	99.58	0.10	0.27	0.01		0.01		0.01	0.00	0.02
50	98.49	1.04	0.11	0.07	0.07	0.09	0.04	0.00		0.09
51	92.51	4.11	0.65	2.32	0.04	0.06	0.03	0.10	0.00	0.18
52	97.18	1.53	0.45	0.60	0.07	0.02	0.08	0.03	0.01	0.04
53	92.60	3.81	1.01	2.23	0.04	0.03	0.07	0.11		0.10
54	98.46	1.04	0.10	0.06	0.12		0.07	0.06	0.01	0.08
55	98.52	1.18	0.05	0.14	0.02				0.01	0.08
56	93.42	2.98	1.93	0.60	0.18	0.27	0.32	0.18	0.01	0.12
57	98.28	1.41	0.04	0.12	0.02	0.01	0.06	0.02	0.01	0.04
58	98.89	0.64	0.12	0.03	0.02	0.03	0.05	0.05	0.00	0.16
59	98.27	1.43	0.10	0.08	0.03	0.02	0.02	0.01	0.00	0.04
60	89.22	9.68	0.26	0.19	0.11	0.15	0.03		0.15	0.21
61	99.24	0.20	0.33	0.01	0.05	0.01	0.01	0.05	0.03	0.08
62	97.28	1.63	0.41	0.05	0.22	0.01	0.06	0.00	0.18	0.14
63	89.60	6.77	2.12	0.69	0.19	0.41	0.04		0.13	0.05
64	98.94	0.84	0.09	0.05	0.05	0.00		0.00	0.01	0.01
	Unweighted percentages									
31	92.69	3.41	1.81	0.70	0.14	0.36	0.22	0.06	0.17	0.45
32	96.61	2.01	0.31	0.37	0.11	0.03	0.11	0.14	0.03	0.28
33	98.94	0.23	0.14	0.34		0.06		0.09		0.20
34	96.40	2.28	0.45	0.08	0.28	0.03	0.11		0.08	0.28
35	95.89	2.37	0.62	0.11	0.34		0.17	0.06	0.17	0.28
36	91.49	3.97	2.13	0.85	0.23	0.43	0.28	0.06	0.11	0.45
37	96.71	2.16	0.14	0.26	0.17	0.09	0.14	0.03	0.09	0.23
38	98.98	0.26	0.23	0.14		0.09		0.03	0.03	0.26
41	96.55	1.78	0.49	0.55	0.02	0.09	0.13		0.08	0.30
42	99.02	0.41	0.09	0.15	0.08		0.06	0.02	0.02	0.15
43	92.80	3.68	1.79	0.34	0.34	0.34	0.28	0.21	0.08	0.13
44	98.66	0.46	0.08	0.08	0.10	0.08	0.11	0.15	0.10	0.19
45	89.94	9.04	0.31	0.12	0.10	0.19	0.04	0.06	0.12	0.10
46	99.32	0.25	0.14	0.06	0.04	0.06				0.14
47	99.08	0.37	0.08	0.06	0.06	0.04	0.04	0.12	0.04	0.12
48	89.65	6.77	1.89	0.67	0.11	0.55	0.04	0.06	0.11	0.13
49	99.28	0.15	0.25	0.06		0.06		0.04	0.06	0.11
50	98.16	0.98	0.13	0.13	0.11	0.06	0.13	0.02		0.27
51	91.18	4.65	0.77	2.59	0.11	0.09	0.13	0.09	0.04	0.34
52	96.42	1.76	0.83	0.55	0.06	0.09	0.11	0.02	0.02	0.13
53	90.54	4.69	1.04	2.79	0.08	0.19	0.11	0.19		0.36
54	97.81	1.21	0.25	0.04	0.23		0.13	0.08	0.08	0.19
55	99.06	0.66	0.04	0.04	0.04				0.02	0.15
56	92.96	3.20	1.86	0.48	0.38	0.25	0.21	0.27	0.08	0.32
57	97.79	1.48	0.17	0.17	0.11	0.02	0.10	0.06	0.02	0.08
58	98.70	0.56	0.10	0.06	0.08	0.08	0.04	0.04	0.04	0.33
59	97.69	1.36	0.19	0.19	0.15	0.08	0.06	0.06	0.02	0.21
60	86.85	11.63	0.46	0.25	0.11	0.10	0.08		0.23	0.30
61	99.08	0.42	0.08	0.04	0.04	0.02	0.06	0.13	0.08	0.06
62	97.19	1.51	0.43	0.17	0.30	0.06	0.09	0.02	0.02	0.21
63	90.05	6.26	1.94	0.70	0.19	0.46	0.08		0.23	0.10
64	98.79	0.85	0.04	0.11	0.04	0.02		0.04	0.04	0.08

Table 12.21 Distribution of not-reached items by financial literacy booklet

	Number of non-reached items									
	0	1	2	3	4	5	6	7	8	>8
Booklet	Weighted percentages									
71	82.17	1.77	1.97	0.21	1.43	0.54	0.67	0.59	0.32	10.33
72	83.34	1.92	1.47	0.71	1.69	0.56	2.03	0.20	0.41	7.70
73	84.04	1.53	1.17	0.89	0.50	1.10	0.04	1.33	1.60	7.81
74	81.41	1.36	0.47	0.38	1.24	0.37	1.34	0.32	0.74	12.38
UH	99.54			0.23			0.23			0.00
	Unweighted percentages									
71	82.50	1.47	2.78	0.29	1.49	0.65	0.81	0.73	0.29	8.99
72	82.51	1.32	1.45	0.99	1.57	0.76	2.64	0.27	0.61	7.89
73	84.42	2.08	1.28	0.94	0.54	1.09	0.12	1.13	1.17	7.24
74	80.87	2.28	0.57	0.74	0.83	0.57	1.89	0.32	0.78	11.16
UH	96.49			1.75			1.75			0.00

BOOKLET EFFECTS

The booklet parameters for the paper-based test that are described in Chapter 9 are reported in Table 12.22. The booklet effects are the amount that must be added to the proficiencies of students who responded to each booklet. That is, a positive value indicates a booklet that was harder than the average while a negative value indicates a booklet that was easier than the average. Since the booklet effects are deviations from an average they sum to zero for each domain. Table 12.23 shows the booklet effects after transformation to the PISA scales.

Table 12.22 Estimated booklet effects in logits

Booklet	Domains		
Standard set	Mathematics	Reading	Science
1	-0.054		0.073
2	0.095	0.179	-0.273
3	0.059	-0.421	0.071
4	-0.064	0.063	
5	0.001		-0.046
6	-0.055	0.120	
7	0.076		-0.063
8	0.127	-0.013	0.052
9	-0.076	0.063	
10	-0.005		-0.032
11	-0.038	-0.121	
12	0.058	0.063	0.130
13	-0.123	0.068	0.089
Easy set			
21	-0.072		0.120
22	0.216	0.271	-0.257
23	0.111	-0.382	0.050
24	-0.075	-0.028	
25	-0.004		-0.171
26	-0.006	0.099	
27	0.147		-0.137
8	0.017	-0.124	0.036
9	-0.105	0.050	
10	-0.072		0.077
11	-0.139	0.011	
12	0.019	-0.109	0.196
13	-0.035	0.212	0.088

Table 12.23 Estimated booklet effects on the PISA scale

Booklet	Domains		
Standard set	**Mathematics**	**Reading**	**Science**
1	-4.2		6.8
2	7.4	14.3	-25.5
3	4.6	-33.8	6.6
4	-5.0	5.0	
5	0.1		-4.3
6	-4.3	9.6	
7	5.9		-5.9
8	9.9	-1.0	4.8
9	-5.9	5.0	
10	-0.4		-3.0
11	-2.9	-9.7	
12	4.5	5.0	12.1
13	-9.6	5.4	8.3
Easy set			
21	-5.6		11.2
22	16.8	21.7	-24.0
23	8.6	-30.7	4.6
24	-5.8	-2.3	
25	-0.3		-15.9
26	-0.5	8.0	
27	11.4		-12.8
8	1.3	-10.0	3.3
9	-8.2	4.0	
10	-5.6		7.1
11	-10.8	0.9	
12	1.4	-8.8	18.3
13	-2.8	17.0	8.2

Booklets that include a single domain cluster at the beginning of the booklet (reading in booklets 3 and 23 and science in booklets 2 and 22) have the largest negative parameters. Booklets with a single domain cluster at the end of the booklet have the highest positive parameters. The mathematics booklet effects for the non-standard easier set of booklets are bigger than for the standard set of booklets.

Estimated booklet parameters for the financial literacy assessment are included in Tables 12.24 and 12.25.

Table 12.24 Estimated booklet effects in logits for financial literacy

	Domains		
Booklet	**Mathematics**	**Reading**	**Financial literacy**
71	0.042	0.475	-0.166
72	0.358	0.124	-0.169
73	-0.197	-0.224	0.171
74	-0.203	-0.375	0.165

Table 12.25 Estimated booklet effects on the PISA scale for financial literacy

	Domains		
Booklet	**Mathematics**	**Reading**	**Financial literacy**
71	3.2	39.6	-13.9
72	27.2	10.3	-14.1
73	-15.0	-18.6	14.2
74	-15.4	-31.2	13.7

Booklets that include financial literacy, with the domain clusters at the beginning of the booklet (mathematics and reading in booklets 73 and 74, and with the financial literacy clusters at the beginning in booklets 71 and 72) have negative parameters. Booklets with the domains at the end of the booklet (mathematics and reading in booklets 71 and 72, and financial literacy in booklets 73 and 74) have positive parameters.

After scaling the PISA 2012 data for each country separately, the booklet parameters were added to the students' achievement scores for mathematics, reading and science.

The mean performance scores could be compared across countries and across booklets.

Tables 12.26, 12.27 and 12.28 present the results of testing the variance in booklet means by country (UH booklet excluded), in each domain. The table rows represent countries and the columns booklets, the cells contain the mean performance by booklet and the squared difference between the observed and expected mean, divided by the error variance by booklet. The expected mean is the average of the booklet means, each weighted by the reciprocal of their error variance. The sum of the squared differences divided by their error variance is chi-square distributed with 13–1=12 degrees of freedom (where 13 represent the number of booklets). Significant values are in bold.

Taking the square root of the squared difference between the observed and the expected mean, divided by the error variance, gives a *z*-score and is an indication of the magnitude of the difference between the observed booklet mean and the expected booklet mean. Significantly easier- or harder-than-expected booklets are bold. Shaded columns are booklets without items in the domain.

The booklet means for domains that are not included in the booklet (shaded columns for reading in Table 12.27 and science in Table 12.28) do not significantly differ from the expected booklet means for all countries, which is to be expected using the deviation contrast codes for booklets in the conditioning model.

There is no significant booklet effect at the international level for the standard and easy booklet sets, because the booklet corrections controlled for this effect.

Table 12.29 presents the results of testing the variance in financial literacy booklet means by country (UH booklet excluded).

Estimation of the booklet effect for the computer-based assessment was not necessary because the length of the computer-based assessment is 40 minutes compared to the two hour paper-based assessment.

Overview of the PISA cognitive reporting scales

PISA 2012 is the fifth PISA assessment and also the fifth occasion on which reading, mathematics and science literacy scores have been reported. Table 12.30 provides a listing of the 28 distinct cognitive scales that have been produced as part of PISA 2000, 2003, 2006, 2009 and 2012.[6] The cognitive scales can be classified into four types: PISA literacy scales, PISA literacy subscales, special purpose scales and PISA computer-based scales.

In the table each scale is named, the database upon which it was established is given, the datasets for which it is provided are indicated (a "P" indicates that the dataset exists, and "(M)" indicates that the domain was a major domain in the cycle); and comments are made about the scale's appropriate use. A detailed overview of each PISA cognitive reporting scale established prior to PISA 2012 has been provided in the *PISA 2009 Technical Report* (OECD, 2011).

[Part 1/2]

Table 12.26 Estimated booklet effects on the PISA scale for mathematics

		Expected mean	Booklet 1 or 21		Booklet 2 or 22		Booklet 3 or 23		Booklet 4 or 24		Booklet 5 or 25		Booklet 6 or 26		Booklet 7 or 27	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	**Australia**	504	493	**8.39**	495	**6.02**	501	1.01	498	2.63	505	0.02	508	1.50	503	0.17
	Austria	508	505	0.42	520	**6.21**	509	0.01	506	0.12	516	3.42	507	0.05	517	3.52
	Belgium	519	522	0.71	520	0.08	509	**10.73**	520	0.05	521	0.26	521	0.42	517	0.43
	Canada	518	515	0.47	529	**10.85**	517	0.15	514	1.60	521	0.61	516	0.22	522	1.41
	Chile	423	419	0.46	428	1.33	423	0.00	416	2.13	422	0.02	423	0.02	422	0.02
	Czech Republic	505	497	2.05	515	**3.88**	502	0.29	510	1.12	510	1.36	504	0.05	500	0.58
	Denmark	502	498	0.64	494	2.67	499	0.35	505	0.54	496	1.94	498	0.95	502	0.00
	Estonia	520	516	0.74	532	**5.47**	522	0.12	513	2.41	532	**5.24**	522	0.14	519	0.07
	Finland	520	514	3.26	525	1.22	526	1.84	509	**7.00**	519	0.15	521	0.10	526	1.93
	France	495	501	1.60	489	1.76	478	**13.98**	487	2.79	495	0.00	497	0.23	489	2.09
	Germany	518	514	0.88	534	**8.63**	515	0.41	516	0.26	521	0.48	514	0.74	510	1.90
	Greece	453	457	0.58	452	0.08	460	2.45	449	0.64	453	0.01	452	0.11	454	0.05
	Hungary	477	459	**12.01**	476	0.01	478	0.05	476	0.01	480	0.31	486	3.17	484	1.63
	Iceland	493	485	1.59	483	1.94	491	0.08	487	0.83	499	1.08	490	0.28	500	1.84
	Ireland	502	483	**14.23**	512	3.84	506	1.13	504	0.18	507	1.53	494	3.22	502	0.01
	Israel	467	472	0.92	474	1.15	459	1.25	469	0.08	471	0.51	462	0.46	462	0.55
	Italy	486	484	0.23	488	0.46	480	3.55	482	1.03	489	1.08	486	0.01	486	0.00
	Japan	537	554	**11.87**	542	1.19	518	**11.46**	536	0.00	539	0.19	541	1.01	533	0.51
	Korea	554	572	**8.41**	557	0.26	544	2.90	552	0.05	574	**9.82**	554	0.00	544	2.22
	Luxembourg	490	484	1.44	499	3.39	475	8.52	488	0.21	496	1.44	493	0.48	481	2.85
	Mexico	413	407	**8.91**	415	0.69	412	0.09	411	0.55	416	1.36	412	0.16	414	0.19
	Netherlands	528	539	**4.09**	516	**4.36**	532	0.77	516	2.63	530	0.26	534	1.76	521	1.55
	New Zealand	499	496	0.68	496	0.30	511	**4.56**	491	2.39	500	0.01	504	0.71	500	0.02
	Norway	489	481	2.10	497	1.43	493	0.33	492	0.46	492	0.28	493	0.66	480	3.28
	Poland	518	518	0.00	523	0.85	520	0.25	513	0.89	513	0.85	520	0.27	501	**7.13**
	Portugal	487	496	1.86	488	0.01	480	1.25	469	**7.76**	485	0.07	479	1.76	482	0.90
	Slovak Republic	482	478	0.51	483	0.03	472	3.40	485	0.22	489	1.48	491	3.63	482	0.00
	Slovenia	503	499	0.78	494	2.16	492	3.61	496	1.09	512	3.14	504	0.01	504	0.01
	Spain	484	484	0.03	490	1.37	490	2.37	482	0.21	486	0.28	483	0.06	477	**4.39**
	Sweden	478	476	0.15	472	1.27	476	0.12	469	3.25	483	0.75	483	0.69	487	3.31
	Switzerland	531	530	0.03	534	0.21	529	0.24	533	0.11	539	3.84	532	0.02	535	0.57
	Turkey	447	439	1.84	418	**27.18**	462	**5.72**	443	0.66	435	**5.43**	444	0.30	455	1.52
	United Kingdom	494	493	0.06	498	0.83	492	0.14	496	0.06	496	0.08	496	0.09	487	2.02
	United States	481	478	0.17	487	1.28	488	1.63	490	3.39	487	1.64	483	0.23	481	0.00
	United States (Connecticut)	504	507	0.03	515	0.76	498	0.27	501	0.09	502	0.10	516	1.65	529	**4.85**
	United States (Florida)	466	466	0.00	457	1.12	456	1.16	484	2.68	483	**4.48**	462	0.29	459	0.96
	United States (Massachusetts)	514	504	0.75	529	3.19	529	1.61	512	0.03	519	0.32	510	0.07	516	0.07
Partners	**Albania**	396	386	2.61	383	2.74	393	0.23	393	0.17	393	0.22	394	0.14	404	2.27
	Argentina	390	374	**8.05**	376	**6.57**	389	0.01	381	1.80	388	0.37	399	3.72	394	0.98
	Brazil	389	385	1.71	380	**7.36**	384	2.07	384	2.69	383	3.82	395	**4.40**	387	0.45
	Bulgaria	439	449	3.68	442	0.24	446	1.34	442	0.48	436	0.23	447	3.32	438	0.01
	Colombia	377	371	1.68	383	1.58	362	**8.52**	380	0.36	378	0.11	374	0.48	381	0.86
	Costa Rica	407	399	2.85	415	2.05	401	1.48	398	3.64	399	2.24	401	1.08	416	**4.16**
	Croatia	470	473	0.17	472	0.09	467	0.41	471	0.04	477	1.78	466	1.02	479	3.66
	Hong Kong-China	561	554	1.87	572	**4.77**	567	1.08	571	3.08	566	0.62	553	1.70	566	1.16
	Indonesia	374	383	2.03	357	**11.99**	362	**5.77**	368	1.71	370	0.78	369	1.07	367	1.99
	Jordan	385	386	0.01	371	**9.09**	398	**5.52**	387	0.25	381	1.04	379	1.98	383	0.40
	Kazakhstan	432	436	0.62	433	0.02	437	0.92	431	0.04	431	0.04	443	**7.85**	430	0.34
	Latvia	491	493	0.12	483	1.76	500	2.57	488	0.33	492	0.10	492	0.05	494	0.25
	Liechtenstein	535	535	0.00	532	0.02	542	0.10	548	0.39	531	0.05	527	0.24	539	0.03
	Lithuania	479	473	0.93	472	1.59	488	2.55	476	0.26	483	0.78	483	0.68	475	0.62
	Macao-China	538	539	0.04	548	**3.86**	541	0.53	544	1.21	534	0.62	539	0.06	540	0.25
	Malaysia	421	412	**3.96**	408	**7.46**	412	3.15	421	0.00	421	0.01	411	**4.72**	412	2.79
	Montenegro	410	414	0.85	395	**8.40**	404	1.54	408	0.14	412	0.12	409	0.06	410	0.00
	Peru	368	356	**5.62**	372	0.53	359	3.59	363	0.69	370	0.05	367	0.07	372	0.88
	Qatar	376	390	**14.10**	369	**5.10**	367	**5.39**	373	0.64	373	0.86	375	0.08	370	3.62
	Romania	445	442	0.28	447	0.15	445	0.01	443	0.14	452	2.69	444	0.02	439	1.39
	Russian Federation	482	489	2.43	480	0.08	489	2.89	491	2.81	496	**11.29**	486	0.98	466	**13.17**
	Russian Federation (Perm)	484	482	0.01	495	1.47	477	0.71	483	0.00	501	**4.51**	488	0.43	476	0.75
	Serbia	449	459	**3.87**	452	0.27	455	1.26	455	1.45	448	0.05	449	0.02	444	0.83
	Shanghai-China	612	619	1.57	632	**16.38**	622	**4.94**	621	3.06	617	1.04	610	0.27	616	0.63
	Singapore	573	573	0.02	584	2.86	583	2.78	580	1.32	568	0.90	572	0.01	570	0.40
	Chinese Taipei	560	561	0.07	562	0.10	566	1.56	568	2.00	560	0.01	542	**9.46**	563	0.23
	Thailand	428	417	**5.85**	396	**35.74**	429	0.13	421	1.90	420	2.52	432	1.14	439	**6.69**
	Tunisia	387	389	0.11	403	**7.96**	382	1.08	398	**4.02**	391	0.47	382	1.36	380	1.93
	United Arab Emirates	434	441	3.01	442	**3.97**	432	0.27	440	2.43	435	0.12	426	**4.07**	437	0.37
	Uruguay	409	416	1.94	400	2.65	402	3.16	402	3.31	403	1.55	411	0.13	408	0.07
	Viet Nam	511	518	1.18	505	0.93	532	**10.62**	528	**6.80**	510	0.02	509	0.07	520	2.89
Booklet set	Standard	494	493	0.49	494	0.05	493	0.40	493	1.29	496	**8.64**	493	0.02	493	0.30
	Easy	416	415	0.23	416	0.05	416	0.00	416	0.02	415	0.61	416	0.01	416	0.04

[Part 2/2]

Table 12.26 Estimated booklet effects on the PISA scale for mathematics

		Expected mean	Booklet 8		Booklet 9		Booklet 10		Booklet 11		Booklet 12		Booklet 13		Chi-sq
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	(df=12)
OECD	**Australia**	504	500	1.60	503	0.09	515	**8.32**	504	0.00	512	**5.60**	519	**13.68**	**49.0**
	Austria	508	503	1.20	491	**11.14**	507	0.02	505	0.54	523	**9.86**	495	**9.29**	**45.8**
	Belgium	519	522	0.53	516	0.66	521	0.32	529	**5.55**	519	0.02	517	0.36	20.1
	Canada	518	517	0.19	513	2.43	520	0.31	512	3.14	519	0.10	521	0.55	**22.0**
	Chile	423	424	0.05	422	0.05	420	0.34	421	0.12	426	0.49	430	2.14	7.2
	Czech Republic	505	512	2.35	508	0.42	502	0.39	487	**10.05**	505	0.00	503	0.04	**22.6**
	Denmark	502	504	0.15	507	1.33	500	0.25	504	0.20	507	1.46	510	2.73	13.2
	Estonia	520	516	1.25	522	0.14	517	0.35	512	3.25	523	0.35	522	0.12	19.7
	Finland	520	515	1.00	517	0.64	527	2.98	514	2.31	528	2.65	524	1.04	**26.1**
	France	495	494	0.06	506	**5.46**	501	1.64	505	3.71	491	0.54	504	3.84	**37.7**
	Germany	518	509	3.68	510	3.13	522	0.40	516	0.31	539	**19.59**	517	0.09	**40.5**
	Greece	453	448	0.81	455	0.08	452	0.12	454	0.06	447	1.47	455	0.10	6.6
	Hungary	477	482	0.82	477	0.01	481	0.72	466	**5.61**	476	0.02	483	1.06	**25.4**
	Iceland	493	497	0.44	483	2.99	503	3.21	492	0.04	496	0.30	499	0.96	15.6
	Ireland	502	503	0.04	494	2.63	507	1.70	497	0.87	500	0.15	510	3.68	**33.2**
	Israel	467	469	0.11	470	0.28	466	0.01	475	1.38	453	**4.31**	460	0.76	11.8
	Italy	486	486	0.07	486	0.00	487	0.40	484	0.36	491	2.73	483	0.56	10.5
	Japan	537	516	**17.68**	542	1.31	534	0.53	545	2.58	529	2.25	545	3.01	**53.6**
	Korea	554	542	3.44	549	0.75	550	0.40	556	0.20	542	3.67	564	3.28	**35.4**
	Luxembourg	490	482	2.44	488	0.29	488	0.25	494	0.66	503	**5.12**	498	3.21	**30.3**
	Mexico	413	413	0.01	409	2.75	411	0.90	408	**6.66**	419	**9.59**	426	**23.52**	**55.4**
	Netherlands	528	519	3.70	526	0.09	524	0.51	535	2.35	527	0.05	533	1.08	**23.2**
	New Zealand	499	494	1.03	504	0.67	502	0.11	486	**5.67**	502	0.21	510	2.83	19.2
	Norway	489	476	**5.35**	493	0.39	498	2.53	485	0.43	490	0.04	490	0.05	17.3
	Poland	518	523	0.87	516	0.16	528	3.19	520	0.14	519	0.04	516	0.04	14.7
	Portugal	487	485	0.11	499	**4.06**	488	0.02	494	1.72	496	2.14	490	0.28	**21.9**
	Slovak Republic	482	496	**7.29**	471	**6.41**	476	0.94	480	0.15	487	0.55	477	0.69	**25.3**
	Slovenia	503	512	2.38	496	2.05	512	2.23	500	0.28	503	0.00	516	**4.32**	**22.1**
	Spain	484	476	**5.49**	484	0.01	493	**6.16**	487	0.61	480	0.85	485	0.05	**21.9**
	Sweden	478	477	0.08	479	0.01	473	1.39	476	0.29	485	1.67	482	0.35	13.3
	Switzerland	531	524	**3.84**	530	0.02	519	**4.06**	531	0.01	540	**3.98**	526	1.15	18.1
	Turkey	447	453	0.93	451	0.43	459	3.74	449	0.09	455	1.94	460	**4.11**	**53.9**
	United Kingdom	494	497	0.25	496	0.20	504	**4.14**	484	3.62	490	0.51	492	0.10	12.1
	United States	481	476	0.39	481	0.01	478	0.17	473	1.90	459	**17.30**	488	1.07	**29.2**
	United States (Connecticut)	504	499	0.30	498	0.56	504	0.01	496	0.91	488	2.17	522	1.30	13.0
	United States (Florida)	466	467	0.00	468	0.04	460	0.41	463	0.11	456	1.16	490	**4.92**	17.3
	United States (Massachusetts)	514	504	1.10	516	0.07	517	0.12	508	0.34	491	**4.95**	519	0.20	12.8
Partners	**Albania**	396	415	**17.87**	412	**6.57**	390	0.87	407	2.83	378	**8.31**	377	**9.30**	**54.1**
	Argentina	390	392	0.18	392	0.07	391	0.09	383	1.75	403	**10.36**	387	0.21	**34.2**
	Brazil	389	388	0.25	388	0.16	390	0.10	389	0.00	399	**12.70**	400	**12.93**	**48.6**
	Bulgaria	439	430	2.84	439	0.01	433	1.05	430	2.21	428	**4.43**	442	0.33	20.2
	Colombia	377	384	2.64	372	1.10	374	0.31	375	0.17	385	3.74	377	0.01	**21.6**
	Costa Rica	407	416	3.45	414	2.69	405	0.08	405	0.20	412	1.00	410	0.29	**25.2**
	Croatia	470	486	**8.43**	458	**10.45**	468	0.27	471	0.01	479	2.10	455	**6.05**	**34.5**
	Hong Kong-China	561	563	0.22	561	0.00	564	0.27	559	0.10	557	0.74	543	**13.18**	**28.8**
	Indonesia	374	396	**16.93**	390	**9.87**	385	**5.39**	393	**12.47**	365	**3.88**	369	1.31	**75.2**
	Jordan	385	401	**11.44**	383	0.23	385	0.00	389	0.61	388	0.38	382	0.52	**31.5**
	Kazakhstan	432	423	3.19	438	1.50	434	0.13	424	2.72	418	**9.02**	434	0.17	**26.6**
	Latvia	491	486	0.82	498	2.36	480	3.04	486	0.65	489	0.07	494	0.19	12.3
	Liechtenstein	535	528	0.17	525	0.30	543	0.14	509	1.59	550	0.67	553	0.72	4.4
	Lithuania	479	471	2.34	482	0.44	484	1.68	475	0.61	483	0.92	478	0.01	13.4
	Macao-China	538	543	1.02	534	0.80	542	0.89	532	1.76	536	0.15	524	**7.95**	19.1
	Malaysia	421	451	**49.67**	419	0.19	427	2.16	438	**10.68**	411	**3.93**	423	0.14	**88.9**
	Montenegro	410	415	1.13	409	0.09	413	0.37	414	0.69	418	2.16	406	0.60	16.1
	Peru	368	375	1.94	368	0.00	373	1.11	364	0.77	373	0.84	373	0.57	16.6
	Qatar	376	382	2.11	383	3.08	382	2.17	388	**10.81**	372	1.57	370	3.76	**53.3**
	Romania	445	430	**9.18**	468	**24.57**	444	0.01	453	2.69	433	**6.00**	438	1.71	**48.8**
	Russian Federation	482	480	0.12	476	2.23	489	2.42	474	2.25	464	**14.62**	483	0.07	**55.4**
	Russian Federation (Perm)	484	485	0.03	482	0.04	472	1.83	477	0.54	480	0.13	488	0.17	10.6
	Serbia	449	441	2.08	445	0.54	444	0.80	441	2.60	455	1.76	447	0.18	15.7
	Shanghai-China	612	605	2.11	607	1.09	607	0.76	610	0.19	587	**17.76**	594	**10.97**	**60.8**
	Singapore	573	579	1.82	566	1.39	575	0.14	580	1.37	568	0.62	557	**8.18**	**21.8**
	Chinese Taipei	560	562	0.17	573	**7.56**	567	1.81	560	0.00	554	1.20	539	**17.65**	**41.8**
	Thailand	428	428	0.03	436	**4.58**	438	**4.48**	427	0.02	436	2.75	428	0.04	**65.9**
	Tunisia	387	401	**5.43**	395	1.51	387	0.01	391	0.61	377	**4.71**	366	**14.42**	**43.6**
	United Arab Emirates	434	444	**5.06**	436	0.25	436	0.24	433	0.14	418	**13.97**	420	**9.79**	**43.7**
	Uruguay	409	405	1.23	411	0.11	410	0.01	408	0.15	417	2.78	428	**18.22**	**35.3**
	Viet Nam	511	512	0.02	498	**4.61**	501	2.67	511	0.00	513	0.19	490	**11.54**	**41.5**
Booklet set	Standard	494	494	0.32	494	0.01	495	2.83	492	1.69	493	0.49	493	0.48	17.0
	Easy	416	417	0.77	417	0.41	416	0.08	415	1.41	417	0.47	416	0.08	4.2

[Part 1/2]

Table 12.27 Estimated booklet effects on the PISA scale for reading

		Expected mean	Booklet 1 or 21		Booklet 2 or 22		Booklet 3 or 23		Booklet 4 or 24		Booklet 5 or 25		Booklet 6 or 26		Booklet 7 or 27	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	**Australia**	512	508	1.36	509	0.90	505	**3.90**	507	1.31	511	0.19	508	1.08	513	0.21
	Austria	492	490	0.21	497	1.18	481	**4.58**	491	0.01	495	0.38	495	0.55	496	0.63
	Belgium	514	512	0.20	519	1.88	515	0.05	517	0.58	515	0.08	510	1.38	516	0.31
	Canada	523	525	0.27	523	0.01	513	**8.09**	524	0.11	523	0.01	522	0.09	526	0.57
	Chile	441	443	0.24	449	2.85	443	0.22	437	0.55	441	0.01	446	1.56	440	0.04
	Czech Republic	497	495	0.26	497	0.00	495	0.37	494	0.37	505	2.39	501	0.82	497	0.02
	Denmark	499	498	0.07	497	0.17	471	**30.82**	506	2.66	494	0.81	509	**5.63**	492	1.62
	Estonia	517	509	1.83	527	**4.79**	520	0.52	504	**5.99**	518	0.04	509	2.53	513	0.74
	Finland	525	528	0.42	541	**7.77**	515	3.33	521	0.77	526	0.03	520	1.16	531	0.80
	France	505	507	0.11	516	**4.20**	516	**4.00**	490	**8.35**	499	1.17	508	0.35	506	0.08
	Germany	512	514	0.14	526	**7.56**	512	0.00	505	2.02	514	0.27	504	3.40	508	0.51
	Greece	478	477	0.01	462	**9.17**	490	**4.50**	468	2.73	481	0.28	495	**10.84**	480	0.16
	Hungary	488	483	0.98	484	0.56	482	1.75	495	1.18	492	0.45	493	0.88	492	0.45
	Iceland	483	481	0.11	478	0.51	500	**9.76**	472	2.51	484	0.02	482	0.05	487	0.34
	Ireland	523	521	0.13	524	0.01	519	0.63	524	0.02	523	0.00	515	**3.95**	522	0.08
	Israel	485	486	0.01	477	1.06	491	0.75	486	0.03	480	0.56	487	0.05	486	0.00
	Italy	490	490	0.01	485	2.59	492	0.91	493	1.00	491	0.07	485	2.65	490	0.00
	Japan	538	536	0.19	541	0.35	547	2.28	530	1.96	536	0.07	530	2.16	536	0.16
	Korea	536	540	0.40	541	0.73	534	0.19	531	0.62	534	0.09	532	0.45	537	0.02
	Luxembourg	488	486	0.10	494	1.30	495	1.38	479	3.46	484	0.39	493	1.19	483	0.72
	Mexico	424	423	0.35	422	0.99	424	0.02	419	**5.50**	423	0.12	426	0.50	421	0.92
	Netherlands	516	516	0.00	515	0.03	517	0.02	511	0.69	517	0.01	502	**7.11**	515	0.03
	New Zealand	512	508	0.39	514	0.20	519	1.52	511	0.01	513	0.07	505	1.60	507	0.62
	Norway	504	497	0.83	504	0.02	511	1.25	517	3.47	506	0.15	502	0.04	504	0.01
	Poland	518	515	0.36	521	0.19	513	1.18	507	**5.93**	514	0.91	527	**4.49**	510	2.00
	Portugal	488	493	0.69	479	2.74	488	0.00	483	0.59	479	1.99	482	1.24	492	0.71
	Slovak Republic	466	467	0.01	462	0.47	464	0.08	472	1.08	465	0.02	469	0.51	467	0.03
	Slovenia	484	481	0.17	490	1.40	473	2.82	478	0.71	487	0.48	479	0.74	486	0.18
	Spain	488	494	3.05	488	0.01	487	0.03	491	0.96	486	0.26	489	0.11	487	0.04
	Sweden	484	480	0.31	483	0.00	487	0.28	476	1.16	487	0.30	491	1.32	490	1.09
	Switzerland	509	510	0.06	510	0.06	515	2.09	514	1.40	513	0.62	507	0.33	511	0.12
	Turkey	476	477	0.07	468	2.17	451	**19.81**	479	0.42	470	1.10	484	2.02	476	0.00
	United Kingdom	500	501	0.06	503	0.70	490	3.24	497	0.12	493	1.03	503	0.49	499	0.01
	United States	497	500	0.28	506	2.07	488	2.73	500	0.29	499	0.09	507	3.00	502	0.45
	United States (Connecticut)	522	525	0.08	524	0.06	489	**7.78**	512	0.58	520	0.05	539	3.00	537	1.47
	United States (Florida)	493	493	0.00	488	0.34	462	**12.11**	502	0.66	492	0.03	504	1.59	486	0.84
	United States (Massachusetts)	527	535	0.49	535	0.70	525	0.03	515	1.76	528	0.01	528	0.02	535	0.74
Partners	**Albania**	396	395	0.01	358	**15.96**	412	3.31	389	0.75	395	0.01	399	0.22	394	0.06
	Argentina	397	396	0.00	403	1.06	410	**4.73**	381	3.41	397	0.02	399	0.17	391	0.91
	Brazil	406	405	0.15	398	**4.89**	418	**10.42**	401	2.80	406	0.04	405	0.36	408	0.11
	Bulgaria	436	436	0.00	434	0.10	444	0.82	444	0.83	439	0.08	447	2.30	434	0.08
	Colombia	403	407	0.47	396	1.87	405	0.13	414	**4.71**	406	0.36	402	0.09	404	0.01
	Costa Rica	441	446	0.96	438	0.17	453	**6.46**	433	2.05	436	1.06	438	0.30	440	0.02
	Croatia	484	488	0.38	485	0.02	484	0.01	485	0.03	489	0.75	484	0.00	488	0.78
	Hong Kong-China	544	546	0.10	549	0.79	543	0.12	545	0.00	549	0.63	539	1.06	542	0.22
	Indonesia	396	397	0.04	385	**5.40**	385	**4.12**	399	0.35	394	0.17	406	2.55	398	0.09
	Jordan	400	397	0.24	392	1.69	396	0.30	399	0.03	397	0.31	402	0.32	399	0.01
	Kazakhstan	393	393	0.01	400	2.98	368	**24.54**	405	**6.67**	393	0.01	385	**4.98**	387	2.16
	Latvia	489	487	0.07	502	**5.45**	482	1.62	498	2.74	485	0.47	498	2.77	499	3.35
	Liechtenstein	517	514	0.02	519	0.02	532	0.69	526	0.36	512	0.07	498	0.98	504	0.35
	Lithuania	477	478	0.00	485	2.53	486	3.01	468	**4.99**	482	1.12	488	**6.02**	475	0.23
	Macao-China	509	510	0.05	513	0.58	508	0.07	508	0.14	504	0.80	515	2.15	509	0.00
	Malaysia	399	396	0.27	394	0.99	396	0.28	408	2.75	396	0.17	384	**6.03**	394	0.78
	Montenegro	423	429	0.99	416	1.68	417	1.24	415	1.48	423	0.01	434	**4.15**	420	0.34
	Peru	385	383	0.04	377	1.48	378	1.06	394	2.30	384	0.00	384	0.03	385	0.02
	Qatar	388	387	0.03	372	**17.75**	387	0.05	392	1.16	384	0.79	393	1.62	388	0.00
	Romania	437	439	0.04	450	**6.34**	454	**8.03**	425	**5.60**	442	0.61	433	0.70	433	0.56
	Russian Federation	476	474	0.17	471	0.73	492	**17.53**	485	3.32	482	1.95	462	**8.92**	478	0.18
	Russian Federation (Perm)	482	476	0.36	477	0.38	498	2.74	495	1.56	491	0.94	476	0.55	482	0.00
	Serbia	447	449	0.31	453	0.85	437	3.15	442	0.77	447	0.02	459	**5.51**	447	0.00
	Shanghai-China	570	571	0.02	580	**6.72**	562	3.66	568	0.13	570	0.00	565	1.49	571	0.05
	Singapore	540	542	0.16	542	0.06	549	2.44	547	1.06	542	0.08	528	**8.56**	541	0.01
	Chinese Taipei	524	524	0.02	520	0.58	517	1.96	537	**7.40**	523	0.00	508	**9.88**	527	0.35
	Thailand	441	446	1.00	429	**7.77**	430	**7.56**	456	**10.70**	443	0.20	446	1.33	443	0.10
	Tunisia	404	404	0.00	390	**6.45**	409	0.63	417	**4.06**	402	0.08	399	0.98	401	0.27
	United Arab Emirates	441	446	1.28	449	3.72	420	**28.87**	441	0.01	443	0.14	451	**5.04**	440	0.12
	Uruguay	411	418	1.58	405	1.36	421	2.77	403	2.36	407	0.84	403	2.47	414	0.29
	Viet Nam	508	510	0.17	505	0.24	495	**5.20**	521	**6.08**	509	0.06	504	0.52	513	1.26
Booklet set	Standard	492	492	0.05	492	0.20	491	0.46	492	0.01	492	0.03	492	0.01	493	0.53
	Easy	423	425	1.25	422	0.51	423	0.00	423	0.00	423	0.02	424	0.12	422	0.45

[Part 2/2]

Table 12.27 Estimated booklet effects on the PISA scale for reading

		Expected mean	Booklet 8		Booklet 9		Booklet 10		Booklet 11		Booklet 12		Booklet 13		Chi-sq
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	(df=12)
OECD	**Australia**	512	509	0.77	512	0.00	514	0.25	519	**5.06**	518	3.67	521	**5.60**	**24.3**
	Austria	492	487	1.32	487	0.87	492	0.00	494	0.16	497	1.15	491	0.02	11.1
	Belgium	514	506	**5.50**	513	0.04	516	0.16	525	**5.01**	512	0.27	513	0.05	15.5
	Canada	523	520	0.94	518	2.00	524	0.05	527	1.31	531	**5.30**	526	0.68	19.4
	Chile	441	433	2.92	438	0.57	442	0.12	450	2.86	431	**4.96**	442	0.14	17.1
	Czech Republic	497	496	0.05	500	0.44	494	0.27	494	0.43	485	**3.87**	507	2.84	12.1
	Denmark	499	502	0.44	507	2.96	499	0.00	495	0.64	497	0.28	513	**6.96**	**53.1**
	Estonia	517	511	2.02	513	0.66	518	0.03	534	**12.12**	520	0.30	524	2.36	**33.9**
	Finland	525	516	3.55	518	3.28	526	0.00	534	2.24	521	0.56	538	**6.46**	**30.4**
	France	505	502	0.45	503	0.14	510	0.74	526	**11.04**	489	**8.06**	501	0.49	**39.2**
	Germany	512	515	0.26	505	2.59	514	0.20	513	0.08	517	1.20	511	0.07	18.3
	Greece	478	502	**22.09**	480	0.16	477	0.06	456	**16.73**	484	1.44	452	**19.68**	**87.9**
	Hungary	488	491	0.29	488	0.00	489	0.09	485	0.44	492	0.49	485	0.42	8.0
	Iceland	483	486	0.12	479	0.46	485	0.07	492	1.85	473	2.70	474	1.70	20.2
	Ireland	523	515	2.83	517	1.28	519	0.76	536	**6.73**	529	0.84	536	**10.96**	**28.2**
	Israel	485	502	**5.35**	481	0.43	485	0.00	482	0.17	487	0.07	473	1.78	10.3
	Italy	490	489	0.04	490	0.00	489	0.17	487	0.81	499	**7.81**	489	0.02	16.1
	Japan	538	522	**10.76**	544	1.70	539	0.06	556	**12.98**	524	**7.81**	553	**9.46**	**49.9**
	Korea	536	532	0.62	538	0.19	537	0.02	542	1.24	522	**6.58**	546	**3.97**	15.1
	Luxembourg	488	491	0.39	490	0.17	487	0.01	496	2.54	484	0.33	479	2.56	14.5
	Mexico	424	439	**48.89**	428	3.32	422	0.31	416	**11.58**	425	0.17	418	**4.71**	**77.4**
	Netherlands	516	509	2.85	518	0.09	517	0.00	525	2.72	527	**3.98**	524	1.88	19.4
	New Zealand	512	506	1.31	509	0.34	512	0.00	510	0.10	517	0.74	530	**8.06**	15.0
	Norway	504	493	3.07	497	1.13	505	0.06	503	0.00	509	0.84	501	0.21	11.1
	Poland	518	525	1.36	515	0.57	528	2.72	526	1.72	526	1.85	516	0.10	**23.4**
	Portugal	488	484	0.46	495	1.80	486	0.05	487	0.02	503	**6.37**	490	0.14	16.8
	Slovak Republic	466	476	**4.84**	464	0.14	460	0.61	469	0.17	455	3.30	455	2.73	14.0
	Slovenia	484	486	0.29	484	0.02	483	0.03	483	0.00	479	0.41	491	1.57	8.8
	Spain	488	481	**5.09**	482	2.37	491	0.52	494	2.20	491	0.57	485	0.46	15.7
	Sweden	484	488	0.55	483	0.00	481	0.15	479	0.50	482	0.05	475	1.61	7.3
	Switzerland	509	508	0.23	507	0.31	503	1.37	513	0.47	498	**5.05**	509	0.02	12.1
	Turkey	476	488	**6.29**	479	0.36	475	0.00	477	0.06	486	**4.05**	471	0.63	**37.0**
	United Kingdom	500	500	0.01	501	0.08	501	0.06	495	0.67	503	0.55	504	0.52	7.5
	United States	497	493	0.30	504	2.41	487	2.27	488	2.33	489	2.26	494	0.13	18.6
	United States (Connecticut)	522	533	1.54	525	0.12	518	0.18	514	0.65	514	0.46	526	0.13	16.1
	United States (Florida)	493	514	**8.43**	500	0.76	485	0.76	478	1.97	490	0.08	503	0.92	**28.5**
	United States (Massachusetts)	527	531	0.20	531	0.17	525	0.04	527	0.00	515	1.66	522	0.18	6.0
Partners	**Albania**	396	414	**10.67**	401	0.60	393	0.13	386	1.30	392	0.25	388	1.39	**34.7**
	Argentina	397	408	**4.47**	400	0.39	392	0.49	395	0.06	385	**6.07**	393	0.38	**22.2**
	Brazil	406	409	0.70	405	0.38	409	0.66	412	2.36	407	0.02	403	0.53	**23.4**
	Bulgaria	436	430	0.98	436	0.00	432	0.30	416	**6.17**	431	0.54	447	2.21	14.4
	Colombia	403	411	2.26	398	1.55	399	0.74	392	3.48	417	**7.24**	387	**9.31**	**32.2**
	Costa Rica	441	447	1.77	436	0.89	439	0.19	448	1.35	439	0.11	436	0.74	16.1
	Croatia	484	489	1.11	477	**4.37**	485	0.05	473	3.78	488	0.47	482	0.19	11.9
	Hong Kong-China	544	532	**8.11**	541	0.89	546	0.10	560	**10.44**	540	1.16	551	1.98	**25.6**
	Indonesia	396	402	1.42	395	0.01	395	0.02	393	0.34	413	**13.75**	386	3.75	**32.0**
	Jordan	400	413	**5.11**	405	1.40	401	0.04	386	**5.27**	405	1.38	396	0.26	16.4
	Kazakhstan	393	382	**6.96**	398	1.47	390	0.71	397	0.54	399	1.92	409	**14.17**	**67.1**
	Latvia	489	491	0.14	484	0.87	486	0.31	483	0.70	480	1.68	475	**5.72**	**25.9**
	Liechtenstein	517	519	0.02	512	0.08	548	2.14	490	1.56	506	0.32	533	0.71	7.3
	Lithuania	477	476	0.15	473	0.75	477	0.00	482	1.40	465	**7.14**	472	1.24	**28.6**
	Macao-China	509	512	0.32	506	0.66	511	0.13	511	0.30	501	3.07	508	0.14	8.4
	Malaysia	399	384	**9.97**	402	0.57	393	1.20	415	**10.89**	414	**11.65**	400	0.11	**45.7**
	Montenegro	423	430	1.69	435	**6.02**	422	0.06	404	**9.93**	428	0.69	417	0.83	**29.1**
	Peru	385	388	0.31	385	0.00	387	0.23	377	1.65	399	**7.10**	372	3.70	17.9
	Qatar	388	406	**21.27**	405	**21.32**	388	0.01	359	**51.87**	403	**17.55**	373	**13.48**	**146.9**
	Romania	437	417	**13.61**	428	3.32	439	0.06	456	**11.04**	431	1.45	444	1.44	**52.8**
	Russian Federation	476	468	2.42	466	**5.52**	478	0.19	477	0.02	471	0.68	476	0.00	**41.6**
	Russian Federation (Perm)	482	459	**6.48**	472	1.31	477	0.36	492	1.11	496	2.18	483	0.00	18.0
	Serbia	447	455	2.66	452	1.48	445	0.12	432	**4.75**	435	**5.45**	447	0.00	**25.1**
	Shanghai-China	570	573	0.48	567	0.45	568	0.19	580	**5.02**	557	**6.71**	573	0.39	**25.3**
	Singapore	540	520	**19.34**	536	0.94	546	0.87	562	**14.08**	542	0.07	552	**4.53**	**52.2**
	Chinese Taipei	524	502	**16.05**	528	1.41	526	0.22	524	0.02	530	1.58	533	**5.22**	**44.7**
	Thailand	441	444	0.52	443	0.13	440	0.03	420	**18.63**	456	**8.68**	441	0.00	**56.6**
	Tunisia	404	400	0.39	404	0.00	401	0.20	396	1.60	424	**13.43**	405	0.01	**28.1**
	United Arab Emirates	441	445	0.54	449	3.53	442	0.03	432	**5.24**	439	0.18	444	0.43	**49.1**
	Uruguay	411	413	0.14	409	0.19	412	0.00	420	2.53	423	**5.09**	400	**4.97**	**24.6**
	Viet Nam	508	480	**33.17**	498	3.00	507	0.04	533	**20.16**	499	2.13	531	**19.69**	**91.7**
Booklet set	Standard	492	492	0.05	492	0.10	492	0.06	492	0.01	492	0.03	492	0.02	1.6
	Easy	423	423	0.07	423	0.27	423	0.12	423	0.07	424	0.63	424	0.45	4.0

[Part 1/2]

Table 12.28 **Estimated booklet effects on the PISA scale for science**

		Expected mean	Booklet 1 or 21		Booklet 2 or 22		Booklet 3 or 23		Booklet 4 or 24		Booklet 5 or 25		Booklet 6 or 26		Booklet 7 or 27	
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2
OECD	**Australia**	521	522	0.05	519	0.21	511	**8.31**	518	0.81	515	3.07	523	0.39	521	0.03
	Austria	508	506	0.16	495	**6.75**	519	**4.66**	509	0.04	518	**5.36**	505	0.80	508	0.00
	Belgium	510	517	**3.86**	514	1.01	496	**19.29**	514	0.91	490	**23.82**	512	0.17	514	0.97
	Canada	525	527	0.26	532	3.05	525	0.01	524	0.03	520	2.02	525	0.00	524	0.06
	Chile	444	455	**5.45**	461	**11.90**	433	**5.87**	444	0.00	424	**20.46**	442	0.33	441	0.48
	Czech Republic	514	510	0.52	521	2.32	512	0.15	518	0.76	516	0.14	513	0.05	510	0.32
	Denmark	501	494	1.95	483	**12.47**	510	2.45	501	0.00	507	1.00	497	0.83	495	1.64
	Estonia	541	537	0.85	555	**8.13**	537	0.59	543	0.24	525	**11.44**	545	0.92	535	1.56
	Finland	547	553	1.82	544	0.35	534	**6.47**	540	2.03	544	0.74	545	0.16	563	**9.97**
	France	499	502	0.44	502	0.39	492	2.20	496	0.29	479	**18.28**	499	0.00	514	**10.45**
	Germany	529	526	0.81	535	1.20	537	2.48	529	0.01	536	2.71	524	1.47	516	**6.51**
	Greece	467	464	0.31	482	**6.89**	467	0.01	465	0.09	470	0.35	465	0.14	480	**7.20**
	Hungary	495	486	2.90	489	0.87	505	**5.67**	497	0.23	502	1.80	498	0.36	489	1.00
	Iceland	478	469	2.04	475	0.20	470	1.96	473	0.41	487	1.74	474	0.41	489	3.19
	Ireland	523	515	1.78	508	**8.34**	534	**6.54**	520	0.22	532	**4.00**	520	0.27	516	1.96
	Israel	470	470	0.02	487	**5.40**	459	3.17	475	0.57	468	0.17	465	0.64	480	1.87
	Italy	493	491	0.68	498	1.83	498	2.90	494	0.06	502	9.12	495	0.54	487	**7.71**
	Japan	547	550	0.39	555	2.65	547	0.02	550	0.29	548	0.06	549	0.16	527	**15.62**
	Korea	538	556	**11.07**	552	**7.25**	521	**8.66**	538	0.00	525	**5.94**	537	0.02	531	1.66
	Luxembourg	491	491	0.03	501	3.82	490	0.06	495	0.44	480	**4.05**	495	0.51	484	1.73
	Mexico	415	414	0.02	430	**55.75**	410	**4.61**	414	0.10	408	**10.47**	414	0.04	416	0.47
	Netherlands	526	542	**8.43**	531	0.83	511	**10.83**	521	0.35	514	**6.10**	526	0.01	536	2.69
	New Zealand	516	519	0.41	513	0.22	525	2.79	511	0.65	509	1.06	518	0.12	510	0.95
	Norway	495	475	**9.66**	497	0.06	509	**5.35**	506	3.70	502	1.73	495	0.00	480	**6.20**
	Poland	526	530	0.65	540	**7.69**	514	**5.33**	527	0.13	510	**14.03**	530	0.90	513	**4.47**
	Portugal	489	485	0.70	482	1.96	499	3.34	483	1.32	490	0.02	486	0.36	493	0.43
	Slovak Republic	473	474	0.06	487	**5.70**	468	0.86	474	0.02	475	0.22	476	0.39	455	**11.26**
	Slovenia	516	512	0.77	524	2.06	512	0.61	522	0.71	521	0.84	517	0.08	505	3.22
	Spain	496	492	1.77	503	3.08	499	0.44	501	2.51	497	0.21	499	0.86	493	1.02
	Sweden	485	473	3.49	485	0.03	487	0.10	485	0.01	499	**7.40**	483	0.21	493	2.60
	Switzerland	516	521	0.95	507	2.60	513	0.44	519	0.60	512	1.05	519	0.53	514	0.20
	Turkey	464	467	0.40	455	2.78	460	0.58	465	0.08	456	2.28	461	0.26	468	0.78
	United Kingdom	513	510	0.28	498	**10.79**	517	0.49	511	0.13	523	2.49	515	0.06	520	1.61
	United States	497	500	0.28	487	2.87	503	1.05	506	2.04	494	0.31	496	0.02	509	2.68
	United States (Connecticut)	519	533	1.15	505	1.68	510	0.62	519	0.00	504	2.89	524	0.18	545	5.78
	United States (Florida)	485	490	0.25	472	1.63	471	1.93	500	1.55	491	0.56	483	0.04	479	0.59
	United States (Massachusetts)	528	520	0.48	516	2.36	523	0.22	529	0.02	526	0.04	526	0.02	536	1.50
Partners	**Albania**	403	379	**12.91**	368	**17.40**	425	**13.00**	392	2.90	430	**23.98**	405	0.22	413	2.93
	Argentina	406	401	0.69	415	2.56	404	0.11	405	0.01	403	0.29	407	0.03	395	**5.89**
	Brazil	402	402	0.00	404	0.45	400	0.35	402	0.02	397	2.14	402	0.00	409	**5.37**
	Bulgaria	446	452	0.93	448	0.06	448	0.06	448	0.08	446	0.01	450	0.34	435	3.07
	Colombia	398	395	0.68	414	**8.75**	393	1.14	402	0.82	409	**4.85**	398	0.01	410	**7.18**
	Costa Rica	431	439	3.75	443	**4.58**	435	0.94	426	0.81	426	0.96	432	0.07	419	**4.83**
	Croatia	491	494	0.33	487	0.63	495	0.71	492	0.00	493	0.16	490	0.09	488	0.46
	Hong Kong-China	555	563	3.52	556	0.12	552	0.31	556	0.07	548	1.59	556	0.10	547	**4.15**
	Indonesia	382	381	0.01	367	**10.85**	387	1.45	380	0.11	388	1.26	379	0.32	390	3.49
	Jordan	410	408	0.11	397	**5.45**	415	0.92	409	0.03	409	0.04	407	0.19	413	0.69
	Kazakhstan	425	416	**4.32**	415	**4.28**	431	1.63	426	0.06	436	**6.53**	427	0.16	413	**8.15**
	Latvia	502	503	0.09	510	1.93	502	0.02	508	1.35	496	1.67	508	1.12	507	0.93
	Liechtenstein	526	527	0.00	514	0.38	530	0.03	537	0.31	511	0.76	516	0.28	518	0.21
	Lithuania	495	502	1.53	508	**6.23**	492	0.47	494	0.11	488	1.85	499	0.58	493	0.19
	Macao-China	521	514	2.18	519	0.18	522	0.05	517	0.41	522	0.10	524	0.63	517	0.69
	Malaysia	419	414	1.03	411	2.83	425	1.64	421	0.14	419	0.00	413	1.50	427	2.50
	Montenegro	410	417	2.32	408	0.14	414	0.71	413	0.31	424	**7.04**	407	0.28	407	0.30
	Peru	373	368	1.03	378	0.80	369	0.72	372	0.03	383	3.15	372	0.11	386	**9.27**
	Qatar	383	369	**12.51**	372	**11.14**	392	**3.96**	382	0.06	393	**5.20**	385	0.19	401	**22.64**
	Romania	439	440	0.08	425	**8.08**	442	0.40	441	0.14	438	0.07	440	0.04	432	1.96
	Russian Federation	486	476	**4.33**	488	0.12	485	0.09	496	2.91	505	**17.05**	491	1.02	482	0.79
	Russian Federation (Perm)	480	472	1.05	496	3.49	476	0.34	488	0.71	498	**5.27**	479	0.02	474	0.54
	Serbia	444	455	**4.02**	451	1.09	447	0.20	447	0.24	436	3.13	445	0.02	436	2.28
	Shanghai-China	580	594	**6.88**	588	3.30	573	3.32	581	0.03	571	3.50	581	0.09	567	**7.22**
	Singapore	552	559	2.21	541	2.95	544	1.56	555	0.23	536	**10.47**	555	0.50	561	**4.36**
	Chinese Taipei	523	527	0.76	508	**13.74**	528	1.44	525	0.20	521	0.57	518	1.65	520	0.58
	Thailand	444	444	0.00	432	**7.38**	446	0.11	447	0.42	451	2.37	441	0.53	461	**16.98**
	Tunisia	398	385	**7.16**	385	**6.36**	402	0.84	396	0.08	407	3.06	400	0.16	418	**14.86**
	United Arab Emirates	448	455	2.38	438	**6.46**	448	0.01	448	0.00	441	3.28	448	0.01	459	**6.35**
	Uruguay	416	421	0.78	433	**11.38**	408	2.38	408	2.13	423	2.11	414	0.14	430	**7.64**
	Viet Nam	528	525	0.21	484	**69.50**	549	**15.12**	530	0.11	538	2.97	528	0.00	523	0.98
Booklet set	Standard	497	497	0.00	497	0.05	498	0.39	499	1.68	497	0.00	498	0.11	498	0.06
	Easy	426	427	0.35	426	0.02	427	0.16	426	0.13	426	0.03	426	0.02	427	0.23

[Part 2/2]

Table 12.28 **Estimated booklet effects on the PISA scale for science**

		Expected mean	Booklet 8		Booklet 9		Booklet 10		Booklet 11		Booklet 12		Booklet 13		Chi-sq
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	(df=12)
OECD	**Australia**	521	520	0.22	523	0.25	536	**11.56**	522	0.02	527	2.36	526	1.22	**28.5**
	Austria	508	509	0.03	503	0.78	498	**5.32**	509	0.02	522	**7.88**	505	0.41	**32.2**
	Belgium	510	512	0.33	512	0.25	521	**6.53**	511	0.07	508	0.19	516	1.65	**59.1**
	Canada	525	522	0.90	522	0.84	535	**4.76**	526	0.11	527	0.19	521	0.76	13.0
	Chile	444	442	0.25	443	0.07	461	**12.34**	444	0.02	441	0.67	451	1.72	**59.6**
	Czech Republic	514	512	0.12	516	0.23	511	0.21	504	2.83	511	0.22	520	1.45	9.3
	Denmark	501	502	0.02	506	1.08	496	0.70	502	0.03	513	**6.11**	509	1.93	**30.2**
	Estonia	541	534	3.25	545	0.67	561	**13.10**	541	0.00	543	0.26	542	0.06	**41.1**
	Finland	547	552	0.93	544	0.37	544	0.26	546	0.07	548	0.02	557	3.57	**26.8**
	France	499	500	0.04	503	0.82	500	0.01	497	0.13	503	0.49	501	0.19	**33.7**
	Germany	529	525	0.66	525	0.71	535	0.92	529	0.00	539	**4.69**	522	2.28	**24.5**
	Greece	467	466	0.05	460	1.37	461	0.69	465	0.14	459	2.53	463	0.37	20.1
	Hungary	495	499	0.74	493	0.17	491	0.43	489	1.45	497	0.14	492	0.22	16.0
	Iceland	478	481	0.18	477	0.06	485	0.74	481	0.15	477	0.02	479	0.04	11.1
	Ireland	523	522	0.04	523	0.03	515	2.28	524	0.11	527	0.50	530	3.24	**29.3**
	Israel	470	470	0.01	473	0.11	464	0.44	474	0.25	464	0.88	466	0.34	13.9
	Italy	493	492	0.16	495	0.40	489	1.54	489	2.26	495	0.39	486	3.17	**30.8**
	Japan	547	544	0.17	546	0.00	553	2.14	548	0.06	559	**7.52**	531	**11.50**	**40.6**
	Korea	538	530	2.79	538	0.00	550	**5.03**	539	0.05	540	0.11	535	0.32	**42.9**
	Luxembourg	491	488	0.35	491	0.02	489	0.18	494	0.24	500	1.61	489	0.24	13.3
	Mexico	415	406	**20.79**	415	0.08	426	**21.89**	417	0.88	410	**5.80**	413	0.38	**121.3**
	Netherlands	526	517	3.50	528	0.13	540	**7.15**	531	1.47	524	0.08	528	0.13	**41.7**
	New Zealand	516	511	0.97	516	0.00	519	0.16	513	0.29	517	0.04	525	1.94	9.6
	Norway	495	499	0.42	493	0.06	493	0.13	491	0.29	502	1.19	487	1.66	**30.4**
	Poland	526	521	0.90	528	0.25	551	**18.66**	530	0.70	527	0.04	521	0.41	**54.2**
	Portugal	489	488	0.14	491	0.11	486	0.39	491	0.06	492	0.27	496	1.01	10.1
	Slovak Republic	473	484	**4.27**	475	0.15	476	0.26	469	0.27	474	0.06	459	**5.11**	**28.6**
	Slovenia	516	518	0.13	516	0.00	532	**6.77**	512	0.46	518	0.10	502	**5.53**	**21.3**
	Spain	496	487	**8.46**	493	0.98	509	**11.35**	499	0.56	493	0.34	489	2.55	**34.1**
	Sweden	485	485	0.02	492	0.93	471	**6.24**	484	0.08	487	0.05	479	1.28	**22.4**
	Switzerland	516	520	1.92	518	0.38	514	0.08	514	0.11	510	1.55	520	0.70	11.1
	Turkey	464	462	0.08	461	0.35	462	0.05	469	1.32	468	0.99	471	2.25	12.2
	United Kingdom	513	524	2.79	518	0.87	500	**5.54**	508	0.74	520	1.50	518	0.53	**27.8**
	United States	497	492	0.64	495	0.19	483	**4.10**	500	0.33	497	0.00	501	0.27	14.8
	United States (Connecticut)	519	520	0.00	511	0.89	517	0.09	523	0.17	521	0.03	538	1.64	15.1
	United States (Florida)	485	493	1.34	483	0.04	464	3.80	487	0.04	495	0.88	498	1.29	13.9
	United States (Massachusetts)	528	525	0.12	534	0.45	526	0.03	528	0.00	529	0.03	540	1.06	6.3
Partners	**Albania**	403	426	**30.52**	398	0.77	355	**46.04**	395	1.18	381	**9.43**	394	1.87	**163.1**
	Argentina	406	407	0.04	409	0.19	408	0.08	403	0.27	419	**8.94**	399	1.79	20.9
	Brazil	402	405	1.49	403	0.28	404	0.34	401	0.04	392	**10.11**	402	0.01	20.6
	Bulgaria	446	442	0.59	447	0.02	446	0.00	445	0.03	445	0.04	452	0.57	5.8
	Colombia	398	387	**4.94**	399	0.04	399	0.04	393	0.94	385	**7.47**	392	1.22	**38.1**
	Costa Rica	431	427	0.68	433	0.20	447	**11.42**	428	0.40	423	2.03	419	**4.61**	**35.3**
	Croatia	491	498	1.53	489	0.33	491	0.00	488	0.24	497	1.20	483	1.80	7.5
	Hong Kong-China	555	548	3.00	559	1.01	562	2.28	549	1.24	564	**4.22**	553	0.07	**21.7**
	Indonesia	382	386	0.82	380	0.17	367	**9.66**	387	0.75	382	0.00	391	**4.03**	**32.9**
	Jordan	410	418	3.43	409	0.05	404	0.94	411	0.03	404	1.23	418	2.59	15.7
	Kazakhstan	425	427	0.40	425	0.00	409	**11.08**	425	0.00	449	**34.19**	422	0.38	**71.2**
	Latvia	502	499	0.24	502	0.00	501	0.01	496	0.73	499	0.23	493	1.98	10.3
	Liechtenstein	526	537	0.64	520	0.13	547	1.20	501	1.54	532	0.12	542	0.57	6.2
	Lithuania	495	484	**6.82**	496	0.01	509	**7.20**	494	0.12	492	0.43	495	0.01	**25.5**
	Macao-China	521	518	0.34	522	0.13	530	**4.35**	519	0.15	518	0.32	525	1.11	10.6
	Malaysia	419	417	0.21	422	0.40	416	0.44	424	0.83	412	2.67	432	**5.40**	19.6
	Montenegro	410	404	1.51	410	0.01	400	**4.11**	408	0.28	411	0.04	407	0.39	17.4
	Peru	373	365	2.63	372	0.12	380	1.50	369	0.87	368	1.21	370	0.37	**21.8**
	Qatar	383	394	**7.44**	384	0.06	361	**33.07**	384	0.03	380	1.10	390	3.14	**100.5**
	Romania	439	443	1.25	440	0.12	440	0.03	440	0.04	441	0.15	443	0.64	13.0
	Russian Federation	486	486	0.00	481	1.66	491	1.04	480	1.12	473	**4.97**	483	0.20	**35.3**
	Russian Federation (Perm)	480	472	1.16	476	0.27	470	1.10	480	0.00	484	0.19	476	0.31	14.5
	Serbia	444	440	0.90	445	0.01	454	2.54	443	0.06	451	1.29	433	**4.20**	20.0
	Shanghai-China	580	580	0.01	580	0.00	598	**12.49**	578	0.10	578	0.09	576	0.73	**37.8**
	Singapore	552	551	0.02	547	0.74	550	0.16	555	0.18	547	1.00	567	**7.38**	**31.8**
	Chinese Taipei	523	521	0.31	525	0.13	535	**7.54**	521	0.46	528	1.49	526	0.74	**29.6**
	Thailand	444	449	1.08	440	1.44	427	**13.44**	442	0.32	442	0.18	450	1.77	**46.0**
	Tunisia	398	409	**4.38**	398	0.00	382	**6.71**	396	0.14	387	**4.90**	407	2.90	**51.5**
	United Arab Emirates	448	453	1.03	449	0.02	443	1.16	444	0.71	440	3.12	463	**10.36**	**34.9**
	Uruguay	416	410	2.54	417	0.00	413	0.27	416	0.03	414	0.18	401	**8.70**	**38.3**
	Viet Nam	528	532	0.64	528	0.00	524	0.44	528	0.00	543	**6.46**	536	2.32	**98.8**
Booklet set	Standard	497	498	0.03	498	0.08	497	0.88	495	**5.11**	498	0.13	498	0.01	8.5
	Easy	426	426	0.70	426	0.00	428	1.17	426	0.49	426	0.02	427	0.14	3.5

Table 12.29 Estimated booklet effects on the PISA financial literacy scale

		Expected mean	Booklet 71		Booklet 72		Booklet 73		Booklet 74		Chi-sq
			Mean	Z^2	Mean	Z^2	Mean	Z^2	Mean	Z^2	(df = 3)
OECD	**Australia**	525	517	**4.95**	524	0.09	533	3.91	530	1.10	**10.1**
	Belgium[1]	541	537	0.68	537	0.48	546	1.01	544	0.31	2.5
	Czech Republic	518	519	0.02	513	1.10	521	0.31	521	0.14	1.6
	Estonia	529	523	1.39	522	1.74	533	0.35	540	3.75	7.2
	France	486	484	0.05	485	0.02	494	1.64	481	0.65	2.4
	Israel	477	479	0.09	485	0.81	465	1.66	477	0.00	2.6
	Italy	466	468	0.57	463	0.98	466	0.01	467	0.10	1.7
	New Zealand	520	527	1.10	508	3.10	525	0.59	520	0.00	4.8
	Poland	510	499	3.86	518	2.56	511	0.02	511	0.00	6.4
	Slovak Republic	473	475	0.08	470	0.24	469	0.29	477	0.53	1.1
	Slovenia	487	478	0.90	483	0.56	486	0.01	498	2.72	4.2
	Spain	484	485	0.03	486	0.09	485	0.02	481	0.36	0.5
	United States	492	476	**5.21**	494	0.12	498	0.69	498	1.00	7.0
Partners	**Colombia**	377	405	**16.74**	407	**17.47**	343	**30.04**	360	**4.11**	**68.4**
	Croatia	480	468	**5.78**	475	0.72	492	**5.62**	485	1.16	**13.3**
	Latvia	500	504	0.45	501	0.05	495	0.90	502	0.13	1.5
	Russian Federation	485	499	**5.32**	491	1.16	486	0.02	467	**11.18**	**17.7**
	Shanghai-China	602	594	3.47	593	**4.83**	611	3.37	615	**7.12**	**18.8**

1. Flemish community only.

Table 12.30 Summary of PISA cognitive reporting scales

Name	Established	2000	2003	2006	2009	2012	Comment
	PISA literacy scales						
Print reading	2000	P(M)	P	P	P(M)	P	Trends can be reported between any of the five cycles, by country or by subgroups within countries
Print mathematics	2003		P(M)	P	P	P(M)	Trends can be reported between 2003, 2006, 2009 and 2012 by country or by subgroups within countries
Print science	2006			P(M)	P	P	Trends can be reported between 2006, 2009 and 2012 by country or by subgroups within countries
Print financial literacy	2012					P	
	PISA literacy subscales						
Reading subscale: Retrieving information	2000	P			P		Trends can be reported between 2000 and 2009 by country or by subgroups within countries
Reading subscale: Interpreting texts	2000	P			P		Trends can be reported between 2000 and 2009 by country or by subgroups within countries
Reading subscale: Reflection and evaluation	2000	P			P		Trends can be reported between 2000 and 2009 by country or by subgroups within countries
Reading subscale: Continuous texts	2009				P		
Reading subscale: Non-Continuous texts	2009				P		
Mathematics subscale: Quantity	2003		P			P	Trends can be reported between 2003 and 2012 by country or by subgroups within countries
Mathematics subscale: Uncertainty and data	2003		P			P	Trends can be reported between 2003 and 2012 by country or by subgroups within countries. Label in 2003 was named uncertainty
Mathematics subscale: Space and Shape	2003	P	P			P	Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends between 2000 and other cycles can be reported for countries, but are not optimal for subgroups within countries. Trends can be reported between 2003 and 2012 by country or by subgroups within countries
Mathematics subscale: Change and Relationships	2003	P	P			P	Established in 2003 and then applied to 2000 with a rescaling (no conditioning). Trends between 2000 and other cycles can be reported for countries, but are not optimal for subgroups within countries. Trends can be reported between 2003 and 2012 by country or by subgroups within countries
Mathematics subscale: Formulating	2012					P	
Mathematics subscale: Employing	2012					P	
Mathematics subscale: Interpreting	2012					P	
Science subscale: Explaining Phenomena scientifically	2006			P			
Science subscale: Identifying scientific Issues	2006			P			
Science subscale: Using scientific Evidence	2006			P			
Science subscale: Physical Systems	2006			P			Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal
Science subscale: Earth and Space systems	2006			P			Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal
Science subscale: Living Systems	2006			P			Limited conditioning implemented permitting unbiased estimation by country and by gender. Results for other subgroups are not optimal
	Special purpose scales						
Interim mathematics	2000	P					
Interim science	2000	P	P				
Science Trend 2003-2006	2006		P	P			Uses items that were common to PISA 2003 and PISA 2006.
Financial literacy mathematics	2012					P	Uses items from one mathematics cluster of print PISA 2012 assessment. Items are recalibrated. Results are not comparable to the main PISA 2012
Financial literacy reading	2012					P	Uses items from one reading cluster of print PISA 2012 assessment. Items are recalibrated. Results are not comparable to the main PISA 2012
	Computer-based scales						
Digital reading	2009				P	P	Trends can be reported between 2009 and 2012 by country or by subgroups within countries
Computer-based mathematics	2012					P	
Problem solving	2012					P	

PISA overall literacy scales

PISA overall literacy scales are the key reporting scales that have been established in the year in which the respective domain was the major domain for the first time, since in that year the framework for the domain was fully developed and the domain was comprehensively assessed. When the overall literacy scale is established the mean of the scale is set at 500 and the standard deviation is set at 100 (for the pooled, equally weighted OECD countries) – for example, 500 on the PISA mathematics scale is the mean achievement of assessed students in OECD countries in 2003.

The intention is that these overall literacy scales will stay in place until the specification of the domain is changed or updated.

PISA literacy subscales

Across the five PISA assessments a total of 18 subscales have been prepared and reported. In PISA 2000, three reading aspect-based subscales were prepared; in PISA 2003, four mathematics content-based subscales were prepared, in 2006 a total of six science subscales were prepared; in PISA 2009 two text format subscales were prepared for reading, and in PISA 2012 three mathematics process-based subscales were prepared.

The subscales are typically prepared only in the year in which a domain is a major domain, since when a domain is a major domain there are sufficient items in each sub-area to support the reporting of the subscales. The one exception to this general practice is mathematics, for which the *space and shape* and *change and relationships* subscales were reported for the PISA 2000 data as well as the PISA 2003 data. These subscales, which were established in 2003 when mathematics was the major domain, could be applied to the 2000 data because only these two areas of mathematics had been assessed in PISA 2000 and sufficient common items were available to support the scaling.

For the 2000 data, the mathematics subscales where prepared using a methodology that permits trend analysis at the national level (or at the level of adjudicated regions), but the subscales are not optimal for analysis at the level of student sub-groups.[7]

For science in PISA 2006, two alternative sets of subscales were prepared. The first was a set of three process-based subscales and the second was a set of three content-based subscales. It is important to note that these are alternative scalings that each rely on the same test items. As such, it is inappropriate to jointly analyse subscales that are selected from the alternative scalings. For example, it would not be meaningful or defensible to correlate or otherwise compare performance on the *physical systems* subscale, with performance on the *using scientific evidence* subscale. Furthermore the content-based subscales can be analysed at the national level (or at the level of adjudicated regions), and can be analysed by gender, but they are not optimal for use at the level of any other student sub-groups, whereas the process-based subscales are suitable in addition for sub-group analyses.[8]

The metric of all of the PISA subscales is set so that subscales within a domain can be compared to each other and with the matching overall PISA reporting scale.[9]

Special purpose scales

There are two types of special purpose scales:

- additional scales that can be used as interim and trend scales prior to the establishment of the related PISA overall literacy scales (three scales developed prior to PISA 2012); and
- additional scales that can be used subsets of the main PISA assessment items that are re-calibrated and scaled separately (two scales developed in PISA 2012).

Prior to PISA 2012, three special purpose scales were developed: interim mathematics, interim science and science trends 2003-2006.

An interim mathematics scale was established and reported in PISA 2000. This scale was prepared to provide an overall mathematics score, and it used all of the mathematics items that were included in the PISA 2000 assessment. This scale was discontinued in 2003 when mathematics was the major domain and the new and more comprehensive PISA overall mathematics literacy scale was established.

An interim science scale was established and reported in PISA 2000. This scale was prepared to provide an overall science score, and it used all of the science items that were included in the PISA 2000 assessment. The PISA 2003 science data were linked to this scale so that the PISA 2003 science results were also reported on this interim science scale. For PISA 2006, this scale was not provided since science was the major domain for the first time and the new and more comprehensive overall PISA science scale was established.

To allow comparisons between science outcomes in 2003 and 2006 a science trend 2003-2006 scale was prepared. This scale is based upon the science items that are common to PISA 2003 and PISA 2006 and can be used to examine trends (on those common items) between 2003 and 2006. The PISA 2003 abilities that are based on the common items can be analysed at the national level (or at the level of adjudicated regions), and can be analysed by gender, but they are not optimal for use at the level of any other student sub-groups. The PISA 2006 abilities, associated with the fully developed overall PISA science scale, can be analysed by national subgroups as well.

Two special scales, financial literacy mathematics and financial literacy reading were included in PISA 2012. Only one cluster of print reading and one cluster of print mathematics items were used for these scales. For both scales, the mean of the scale was set at 500 and the standard deviation was set at 100 (for the pooled, equally weighted OECD countries that participated in financial literacy assessment in PISA 2012).

Computer-based scales

There are three computer-based scales.

A digital reading scale was established in 2009, computer-based mathematics and problem solving scales were developed in 2012. For the problem solving scale the mean of the scale was set at 500 and the standard deviation was set at 100 (for the pooled, equally weighted OECD countries). The computer-based mathematics scale was equated to the paper-based mathematical scale so the results could be compared for two modes of assessments.

TRANSFORMING THE PLAUSIBLE VALUES TO PISA SCALES

For PISA 2012, the reading, mathematics and science results are each reported on the scales that were established when the respective domain was a major domain. Therefore in the case of reading, the results are directly comparable with those that have been reported for PISA 2000, 2003, 2006 and 2009. In the case of mathematics they are directly comparable with the results reported in PISA 2003, 2006 and 2009 and for science they are directly comparable with the results reported in PISA 2006 and 2009.

Mathematics

For mathematics, the PISA 2012 plausible values were equated to the PISA scale by using common item equating.

A shift to align the scales was computed as follows: of the 35 mathematics items that were included in the PISA 2009 Main Survey, 34 were used for PISA 2012 Main Survey assessment.

A shift of 0.07 of a logit was required to align PISA 2009 and PISA 2012 mathematics scales. After applying this shift, the same transformation was used as in PISA 2009.

The resulting transformation required to place logits on the PISA mathematics scale was:

$$\text{PISA 2012 scaled score} = ((L + 0.0981) / 1.2838) \times 100 + 500$$

where L is the logit scale outcome of the 2012 scaling.

For details about equating procedures in 2009, see the *PISA 2009 Technical report* (OECD, 2011).

Reading

For reading, the PISA 2012 plausible values were equated to the PISA scale by using common item equating.

A shift to align the scales was computed as follows: of the 101 reading items that were included in the PISA 2009 Main Survey, 44 were selected for PISA 2012 Main Survey assessment. The average item difficulty of the 44 link items was set to zero in PISA 2012 while it was –0.0274 in PISA 2009. A shift of –0.0274 of a logit was therefore required to align PISA 2009 and PISA 2012 reading scales.

After applying this shift, the transformations required to place logits on the PISA reading scale were as given below. Note that the transformation is done separately by gender, as has been the case since PISA 2003.

For female students:

PISA 2012 scaled score = $((0.8739 \times L - 0.4655) / 1.1002) \times 100 + 500$

For male students:

PISA 2012 scaled score = $((0.8823 \times L - 0.5427) / 1.1002) \times 100 + 500$

For students with missing gender code:

PISA 2012 scaled score = $((0.8830 \times L - 0.5079) / 1.1002) \times 100 + 500$

Science

For science, the PISA 2012 plausible values were equated to the PISA scale by using the common items equating method.

Fifty-three science items that were included in the PISA 2009 Main Survey were used again for the PISA 2012 Main Survey assessment. The average item difficulty of the 53 link items was set to zero in PISA 2009 and PISA 2012.

The transformation required to place logits on the PISA science scales was the same as in PISA 2009:

PISA 2012 scaled score = $((L - 0.1646) / 1.0724) \times 100 + 500$

Computer-based mathematics

The computer-based mathematics scale was equated to the paper-based mathematics scale in PISA 2012.

Average item difficulty for computer-based mathematics item set was 0.7232 logits greater then print mathematics items average when both set of items were scaled jointly.

Therefore the transformation to the plausible values required a shift of 0.7232 as the average item difficulty of the computer-based mathematics items was set to zero.

After applying this shift, the transformations required to place logits on the PISA computer-based mathematics scale were as given below.

PISA 2012 scaled score = $((L + 0.8213) / 1.2838) \times 100 + 500$

Digital reading

For digital reading, the PISA 2012 plausible values were equated to the PISA digital reading scale using common item equating.

A shift to align the scales was computed as follows: of the 29 digital reading items that were included in the PISA 2009 Main Survey, 19 were selected for PISA 2012 Main Survey assessment. The average item difficulty of the 19 link items was set to zero in PISA 2012 while it was –0.0604 in PISA 2009. A shift of –0.0604 of a logit was therefore required to align PISA 2009 and PISA 2012 digital reading scales.

After applying this shift, the transformations required to place logits on the PISA digital reading scale were as given below.

The transformation required to place digital reading logits on the PISA digital reading scale was:

PISA 2012 scaled score = $(((L - 0.5769) / 1.1011) \times 96.3956) + 498.9126$

Problem solving

A new scale for problem solving was established in PISA 2012. Therefore the only transformation to the plausible values was a standardisation to an OECD mean of 500 and standard deviation of 100 (using equally weighted, pooled database).

PISA 2012 scaled score = $(((L - 0.0331) / 1.2321) \times 100) + 500$

Financial literacy

A new scale for financial literacy was established in PISA 2012. Therefore the only transformation to the plausible values was a standardisation to an OECD mean of 500 and standard deviation of 100 (using equally weighted, pooled database).

PISA 2012 scaled score = $((L - 0.2554) / 1.0731) \times 100 + 500$

It should be noted that mathematics and reading items included in financial literacy assessment were calibrated and standardised separately of the main PISA 2012 mathematics and reading. Therefore main PISA 2012 mathematics and reading scales are not directly comparable to the financial literacy mathematics and reading scales.

The transformation to the financial literacy mathematics and reading plausible values was a standardisation to an OECD mean of 500 and standard deviation of 100 (using equally weighted, pooled database).

Financial literacy mathematics

PISA 2012 scaled score = $((L + 0.2684) / 1.3171) \times 100 + 500$

Financial literacy reading

PISA 2012 scaled score = $((L - 0.3686) / 1.1992) \times 100 + 500$

LINK ERROR

Link errors estimated using the methodology discussed in Chapter 9 were computed for the following 21 links:

- PISA Mathematics scales 2003 to 2006, 2003 to 2009, 2003 to 2012, 2006 to 2009, 2006 to 2012 and 2009 to 2012;
- PISA Reading scales 2000 to 2003, 2000 to 2006, 2000 to 2009, 2000 to 2012, 2003 to 2006, 2003 to 2009, 2003 to 2012, 2006 to 2009, 2006 to 2012, and 2009 to 2012;
- PISA Science scale 2006 to 2009, 2006 to 2012 and 2009 to 2012; and
- PISA Digital Reading scale 2009 to 2012.

The results are given in Table 12.31.

Table 12.31 Link error estimates

Link	Link Error on PISA scale
PISA Mathematics scale 2003 to 2006	1.350
PISA Mathematics scale 2003 to 2009	1.990
PISA Mathematics scale 2003 to 2012	1.931
PISA Mathematics scale 2006 to 2009	1.333
PISA Mathematics scale 2006 to 2012	2.084
PISA Mathematics scale 2009 to 2012	2.294
PISA Reading scale 2000 to 2003	5.321
PISA Reading scale 2000 to 2006	4.963
PISA Reading scale 2000 to 2009	4.937
PISA Reading scale 2000 to 2012	5.923
PISA Reading scale 2003 to 2006	4.480
PISA Reading scale 2003 to 2009	4.088
PISA Reading scale 2003 to 2012	5.604
PISA Reading scale 2006 to 2009	4.069
PISA Reading scale 2006 to 2012	5.580
PISA Reading scale 2009 to 2012	2.602
PISA Science scale 2006 to 2009	2.566
PISA Science scale 2006 to 2012	3.512
PISA Science scale 2009 to 2012	2.006
PISA Digital Reading 2009 to 2012	2.480

Table 12.32 **Link error covariances**

Link	Link Covariance on PISA scale
PISA Mathematics scale 2006 with 2009, referred to 2003	2.340
PISA Mathematics scale 2006 with 2012, referred to 2003	0.673
PISA Mathematics scale 2009 with 2012, referred to 2003	0.605
PISA Mathematics scale 2009 with 2012, referred to 2006	-0.064
PISA Reading scale 2003 with 2006, referred to 2000	16.492
PISA Reading scale 2003 with 2009, referred to 2000	22.049
PISA Reading scale 2006 with 2009, referred to 2000	21.700
PISA Reading scale 2006 with 2009, referred to 2003	12.054
PISA Science scale 2009 with 2012, referred to 2006	7.445

References

Adams, R.J. (2005), “Reliability as a Measurement Design Effect”, *Studies in Educational Evaluation*, Vol. 31, pp. 162-172.

OECD (2011), *PISA 2009 Technical Report*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.17879789264167872-en

OECD (2008), *PISA 2006 Technical Report*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/ 9789264048096-en

OECD (2005), *PISA 2003 Technical Report*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/ 9789264010543-en

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

Notes

1. For more information, consult *www.oecd.org/pisa*.

2. For more information, consult *www.oecd.org/pisa*.

3. For more information, consult *www.oecd.org/pisa*.

4. The ‘Xs’ represent a different number of students in each graph.

5. Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

6. Note that this section refers to cognitive scales only. PISA has also produced a wide range of other scales that are affective or behavioural scales.

7. This is because conditioning variables were not used in the construction of the scales for the PISA 2000 data (see OECD [2005], *PISA 2003 Technical Report*).

8. This is because gender was the only conditioning variable used in the construction of the content-bases scales (see OECD [2008], *PISA 2006 Technical Report*).

9. Note, of course, that as mentioned above, comparison across alternative scalings of the same domain are not appropriate.

13

Coding Reliability Studies

A substantial proportion of the PISA 2012 items were open-ended and required coding by trained personnel. It was important therefore that PISA implemented procedures that maximised the validity and consistency (both within and between countries) of this coding. Each country coded items on the basis of coding guides prepared by the Consortium using the design described in Chapter 2. Training sessions to train coders from different countries on the use of the coding guides were held prior to both the Field Trial and the Main Survey.

This chapter describes the outcomes of three aspects of the coding reliability studies undertaken in conjunction with the Main Survey. These are *i)* the consistency analyses undertaken to inform the Technical Advisory Group about levels of coding agreement for each of the items that require coder judgement, *ii)* the consistency analyses to assess within-country coder agreement and *iii)* the international coder review undertaken to examine possibilities of countries' coding bias. The consistency analyses are described in the next section and the analyses undertaken for international coder review are described in subsequent sections.

CONSISTENCY ANALYSES

Similar to previous cycles, the consistency analyses were undertaken in relation to a subset of constructed-response items. In PISA 2012 all constructed-response items were classified into two sets. The majority of constructed-response items were classified as *constructed-response expert* items, indicating that they would need some judgement from the coders and, therefore, would need to be included in the multiple-coding exercise and the subsequent analyses. A small number of constructed-response items was classified as *constructed-response manual*, which required coding by coders but did not require multiple-coding due to fairly simple, straightforward coding instructions for the item in the coding guide. Constructed-response manual items are the ones that on the one hand cannot be automatically coded due to limitations of the data management software *KeyQuest*, but on the other hand do not require an expert judgment. An example of such instruction can have code '1' for π or 3.14 or any other approximation of π, and 0 for any other response. The symbol π cannot be entered in *KeyQuest* and such item would be coded manually. More details about item classification can be found in Annex A, Tables A1.1 to A1.7.

The number of constructed-response expert items varied between domains and also depended on the set of booklets administered by the country (standard or easier). The size of the data available for analysis for each domain depended on the number of constructed-response expert items and whether the test was administered in the country in the major or minor language. The way in which items were allocated to coders for multiple coding depended on whether an item was coded by the country on line or on paper.

PISA 2012 offered seven domains in total. There were four paper-based domains: mathematics, reading, science and financial literacy and three computer-based domains: problem solving, mathematics and reading. Participating countries and economies that have more than one language of instruction administered the test in more than one language, however, if the Consortium expected fewer than 50 students per booklet type for a minor language for a particular domain the locale (country-by-language unit) was exempted from the multiple coding of this domain because the amount of data would be insufficient for analysis. In the Main Survey 76 locales participated in the multiple-coding exercise. Table 13.1 shows which locales participated in multiple coding for which domains and with which options.

In the paper-based assessment there were two groups of countries: those that did standard booklets only (booklets 1-13) and those that did some standard booklets and some non-standard easier booklets (booklets 8-13 and 21-27). There were 17 participants that chose this second option. Both easier and standard booklets contained new and link mathematics items as well as science and reading link items. In addition, there were 18 participants from both groups that administered the financial literacy test (see Chapter 2 for details on the PISA 2012 test design)

In the computer-based assessment there were also two groups of participants. Twelve of them assessed their students in only one computer-based domain, problem solving. In addition, there were 32 participants that assessed their students in three computer-based domains: problem solving, computer-based mathematics and digital reading.

In PISA 2012, eleven participants opted to code constructed-response paper-based items using an online coding system. This system was primarily designed for the coding of the constructed-response computer-based items and was used to code constructed-response computer-based items by all participants administering the PISA 2012 computer-based assessment. Coding of the paper-based items in the online coding system was not compulsory and most of the participants coded constructed-response paper-based items in the paper test booklets and in the specially designed multiple-coding sheets and then entered data into the data management software *KeyQuest*.

[Part 1/2]

Table 13.1 **Participation in multiple coding by domain, locale, option**

		Paper-based domains/options				Computer-based domains	
	Locale (country-by-language unit)	**Mathematics, reading and science**	**Easier booklet**	**Financial literacy**	**On-line coding**	**Problem solving**	**Mathematics and digital reading**
OECD	Australia-English	Y		Y	Y	Y	Y
	Austria-German	Y			Y	Y	Y
	Belgium-Flemish	Y		Y		Y	Y
	Belgium-French	Y				Y	Y
	Canada-English	Y				Y	Y
	Canada-French	Y				Y	Y
	Chile-Spanish	Y	Y			Y	Y
	Czech Republic-Czech	Y		Y		Y	
	Denmark-Danish	Y				Y	Y
	Estonia-Estonian	Y		Y		Y	Y
	Finland-Finnish	Y				Y	
	France-French	Y		Y		Y	Y
	Germany-German	Y				Y	Y
	Greece-Greek	Y					
	Hungary-Hungarian	Y				Y	Y
	Iceland-Icelandic	Y			Y		
	Ireland-English	Y				Y	Y
	Israel-Arabic	Y		Y	Y	Y	Y
	Israel-Hebrew	Y		Y	Y	Y	Y
	Italy-Italian	Y		Y		Y	Y
	Japan-Japanese	Y				Y	Y
	Korea-Korean	Y			Y	Y	Y
	Luxembourg-French	Y					
	Luxembourg-German	Y					
	Mexico-Spanish	Y	Y				
	Netherlands-Dutch	Y			Y	Y	
	New Zealand-English	Y		Y			
	Norway-Norwegian	Y				Y	Y
	Poland-Polish	Y		Y		Y	Y
	Portugal-Portuguese	Y				Y	Y
	Slovak Republic-Slovak	Y		Y		Y	Y
	Slovenia-Slovenian	Y		Y		Y	Y
	Spain-Basque	Y				Y	Y
	Spain-Catalan	Y		Y		Y	Y
	Spain-Spanish	Y		Y		Y	Y
	Sweden-Swedish	Y			Y	Y	Y
	Switzerland-French	Y			Y		
	Switzerland-German	Y			Y		
	Turkey-Turkish	Y				Y	
	United Kingdom-English	Y				Y[3]	
	United States-English	Y		Y		Y	Y
Partners	Albania-Albanian	Y					
	Argentina-Spanish	Y	Y				
	Brazil-Portuguese	Y	Y			Y	Y
	Bulgaria-Bulgarian	Y	Y			Y	
	Colombia-Spanish	Y	Y	Y	Y	Y	Y
	Costa Rica-Spanish	Y	Y				
	Croatia-Croatian	Y		Y		Y	
	Cyprus-English[1,2]	Y	Y			Y	
	Cyprus-Greek[1,2]	Y	Y			Y	
	Hong Kong-China-Chinese	Y				Y	Y
	Indonesia-Indonesian	Y					
	Jordan-Arabic	Y	Y				
	Kazakhstan-Kazakh	Y	Y				
	Kazakhstan-Russian	Y	Y				
	Latvia-Latvian	Y		Y			
	Lithuania-Lithuanian	Y					
	Macao-China-Chinese	Y				Y	Y
	Malaysia-English	Y				Y	
	Malaysia-Malay	Y				Y	
	Montenegro-Montenegrin	Y				Y	
	Peru-Spanish	Y	Y				
	Qatar-Arabic	Y					
	Qatar-English	Y					
	Romania-Romanian	Y	Y				
	Russian Federation-Russian	Y		Y		Y	Y

[Part 2/2]
Table 13.1 Participation in multiple coding by domain, locale, option

Locale (country-by-language unit)	Paper-based domains/options				Computer-based domains	
	Mathematics, reading and science	Easier booklet	Financial literacy	On-line coding	Problem solving	Mathematics and digital reading
Serbia-Serbian	Y	Y			Y	
Shanghai-China-Chinese	Y		Y		Y	Y
Singapore-English	Y				Y	Y
Chinese Taipei-Chinese	Y			Y	Y	Y
Thailand-Thai	Y					
Tunisia-Arabic	Y	Y				
United Arab Emirates-Arabic	Y	Y			Y	Y
United Arab Emirates-English	Y	Y			Y	Y
Uruguay-Spanish	Y	Y		Y	Y	
Viet Nam-Vietnamese	Y	Y				

1. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

3. England only.

As was the case in the previous cycles, for the PISA 2012 Main Survey a subset of constructed-response expert items from the first cluster in each booklet was multiple coded. Given that each item appeared in each cluster, this design provided around a hundred students per item for major languages and, at the same time, ensured that the amount of missing data was minimised (the amount of missing data and non-responses increases towards the end of the booklet). For the paper-based multiple coding for their main test language each National Centre was required to randomly assign 100 booklets of each type that they were using for testing, and for minority languages the requirement was at least 50 booklets of each type. Four coders participated in the multiple coding exercise.

For the computer-based coding for their main test language in each participant the online coding system randomly assigned at least 100 records of each constructed-response expert item for multiple coding, and for minority languages it assigned at least 50 records of such items for multiple coding. The actual number of responses assigned for multiple coding depended on the number of coders involved in the coding of the item and the number of records available for coding. For example, if four coders coded an item in the main test language and there is a sufficient number of records for single and multiple coding then 100 records of this item would be randomly chosen by the system for multiple coding. If there were five coders, then the number of responses allocated for multiple coding would increase to 125 to ensure that each of these responses are coded 4 times and each coder coded 100 responses from the pool, and so on.

All analysis was done by item. Each response was coded by four coders. Only students with four non-missing codes were used for analysis. The statistics were first aggregated by locale-domain and then for each item internationally.

The following notation is used for consistency analysis:

$i=1,....,I$ – items in the domain

$c=1,...,C_i$ – locale that retained the item[1]

$j=1,...,J_{i,c}$ – students in each locale who attended to the item i

$k=1,...,K_{ic}$ – coders in each locale who coded item i during multiple coding exercise in the locale c

$x_{ijk}=0, 1, 2, ...$ – code allocated by coder k to student j when coding item i.

To investigate the level of disagreement between coders, the data collected were used to first compute a coder-item disagreement index R_{ikc}. This index was computed for each coder k and each item i across all records j in the multiple coding exercise within a given country-by-language unit c. The index was computed as an average of the absolute value of the residual multiplied by 100 for readability purposes.

13.1

$$R_{ikc}=\frac{100}{J_{ic}}\sum_{j}\left|x_{ijk}-\frac{1}{4}\sum_{k}x_{ijk}\right|$$

R_{ikc} was then aggregated to compute other indices. A value of $R_{ikc}=0$ shows a perfect agreement among coders for all students responding to the item of a particular language in the country (e.g. shaded cells for item A in Table 13.2).

Each disagreement between coders contributes to an increase of the index. For example, if one coder disagrees by one score with three others, all of whom agree with each other, the residual for this one would be 0.75 and the residual for each of three others would be 0.25. In the example in Table 13.2, coder 201 disagrees by one score with three other coders 20% of the time when coding item B and there are no other cases of disagreement for this item (a fictitious situation). In this case $R_{ikc}=15$ for this coder and for the three other coders it is 5.

On the other hand, if two of the coders disagree with the two others in 20% of the cases and there are no other cases of disagreement (this is another fictitious situation with all residuals being 0.5), then $R_{ikc}=10$ for all coders (shaded cells for item C in Table 13.2).

In a real situation there is always a mix of different combinations of disagreement and the R_{ikc} would look more like shaded cells for items D and E in Table 13.2.

Table 13.2 Fictitious examples of various indices calculated on locale-domain level

Coder ID	Item A	Item B	Item C	Item D	Item E	Q_{kc}
201	0	15	10	9.88	11.82	**9.34**
202	0	5	10	4.45	10.91	**6.07**
203	0	5	10	5.14	10.45	**6.12**
204	0	5	10	5.14	10.45	**6.12**
S_{ic}	**0**	**7.5**	**10**	**6.15**	**10.91**	

For each item in each locale, a locale item reliability index S_{ic} was computed as follows:

13.2

$$S_{ic} = \frac{1}{K_{ic}} \sum_{k} R_{ikc}$$

and the average across all items coded by a particular coder, Q_{kc} was calculated as:

13.3

$$Q_{kc} = \frac{1}{I} \sum_{i} R_{ikc}$$

Examples of some S_{ic} values are shown in the bottom line in Table 13.2 and examples of some Q_{kc} values are shown in the last column in Table 13.2. In this example coder 201 appears less reliable than the three other coders. Coder reliability indices were reported to the countries in the national reports to inform countries of the quality of their coders. This index was not aggregated further.

S_{ic} was further aggregated across all OECD locales that retained the item *i* to form the OECD item reliability index (T_i) for all items except financial literacy and easier mathematics paper-based items. The financial literacy items and the easier mathematics items were aggregated across all locales that retained item *i*.

13.4

$$T_i = \frac{1}{C_i} \sum_{c} S_{ic}$$

The OECD/international item reliability index T_i for each item in the multiple-coding exercise is presented in Table 13.3. As was the case in the previous PISA administrations, the items with $T_i > 7.5$ were considered to have high inconsistency of coding and highlighted in grey. The threshold of 7.5 is a rule of thumb which is based on two cycles of experience of analysing variability of coding data for the Field Trial and the Main Survey. As explained previously it can be interpreted as equivalent to the case when one of the coders disagree with three others 20% of the time while three others agree between themselves. Or two coders disagree with two others 15% of the time. The threshold was accepted as high because it does not appear often in the paper-based domains.

Table 13.3 **OECD/International item reliability indices (Ti)**

Item	Number of locales	Ti	S.E.	Ti_SD	Item	Number of locales	Ti	S.E.	Ti_SD
Computer-based mathematics					**Paper-based mathematics**				
CM015Q03	28	5.10	(0.571)	3.019	PM00FQ01	39	3.46	(0.338)	2.113
CM028Q03	28	1.08	(0.186)	0.986	PM00KQ02	41	0.69	(0.096)	0.614
CM038Q05	28	1.87	(0.231)	1.220	PM155Q01	41	1.17	(0.180)	1.149
CM038Q06	28	4.41	(0.476)	2.517	PM155Q02	41	3.50	(0.332)	2.126
Problem solving					PM155Q03	41	5.65	(0.555)	3.554
CP002Q06	33	4.93	(0.394)	2.264	PM406Q01	41	1.08	(0.165)	1.053
CP018Q05	32	2.16	(0.252)	1.424	PM406Q02	41	2.44	(0.333)	2.130
CP034Q05	33	1.15	(0.145)	0.832	PM446Q02	41	1.24	(0.572)	3.664
CP036Q02	33	1.64	(0.352)	2.021	PM462Q01	41	1.65	(0.213)	1.367
CP036Q03	33	1.30	(0.128)	0.737	PM828Q01	41	6.24	(0.475)	3.044
CP041Q02	33	4.62	(0.470)	2.702	PM903Q01	39	3.71	(0.455)	2.839
Digital reading					PM905Q02	39	1.95	(0.190)	1.185
CR002Q05	28	6.23	(0.589)	3.115	PM906Q02	41	8.28	(0.621)	3.976
CR013Q07	28	4.46	(0.512)	2.710	PM909Q03	41	0.51	(0.096)	0.618
CR014Q01	28	5.73	(0.585)	3.096	PM923Q04	39	1.00	(0.160)	1.001
CR017Q07	28	7.26	(0.740)	3.915	PM936Q02	20	0.84	(0.185)	0.827
CR021Q08	23	9.68	(1.194)	5.726	PM942Q02	20	0.70	(0.285)	1.274
Paper-based reading					PM943Q02	39	0.24	(0.063)	0.394
PR220Q01	41	3.87	(0.350)	2.243	PM948Q03	20	0.31	(0.077)	0.345
PR404Q10A	41	4.13	(0.345)	2.209	PM949Q03	39	2.61	(0.322)	2.009
PR404Q10B	41	5.99	(0.494)	3.166	PM953Q02	39	3.78	(0.322)	2.008
PR406Q01	40	2.41	(0.266)	1.684	PM953Q04	39	2.45	(0.225)	1.403
PR406Q02	41	8.05	(0.725)	4.641	PM954Q02	39	1.19	(0.157)	0.983
PR406Q05	41	2.64	(0.321)	2.055	PM954Q04	39	0.89	(0.162)	1.011
PR412Q08	41	5.53	(0.385)	2.467	PM955Q02	41	2.36	(0.240)	1.539
PR420Q06	41	5.37	(0.473)	3.028	PM955Q03	41	1.51	(0.195)	1.246
PR432Q05	41	2.67	(0.220)	1.406	PM961Q02	20	0.62	(0.165)	0.740
PR437Q07	41	6.27	(0.490)	3.140	PM961Q05	20	9.57	(1.739)	7.778
PR446Q06	41	1.79	(0.298)	1.907	PM991Q02	20	0.93	(0.233)	1.042
PR453Q04	41	7.57	(0.666)	4.265	PM992Q03	41	0.82	(0.141)	0.905
PR453Q06	40	3.63	(0.380)	2.405	PM995Q02	39	0.98	(0.468)	2.921
PR455Q02	41	6.04	(0.504)	3.228	**Science**				
PR456Q02	41	3.11	(0.400)	2.563	PS131Q02	40	3.15	(0.373)	2.357
PR456Q06	41	1.34	(0.140)	0.896	PS131Q04D	39	3.77	(0.335)	2.093
PR466Q02	41	1.99	(0.226)	1.449	PS269Q01	41	2.15	(0.268)	1.716
Financial literacy					PS269Q03D	41	2.62	(0.323)	2.067
PF004Q03	19	1.59	(0.322)	1.405	PS326Q01	41	4.42	(0.408)	2.614
PF024Q02	20	7.05	(1.118)	4.998	PS326Q02	41	4.18	(0.449)	2.872
PF028Q02	20	6.96	(0.760)	3.401	PS408Q03	41	5.99	(0.516)	3.305
PF036Q01	20	4.85	(0.622)	2.781	PS425Q03	41	6.95	(0.676)	4.329
PF051Q01	20	2.29	(0.394)	1.760	PS425Q04	41	3.31	(0.433)	2.775
PF051Q02	20	7.60	(1.110)	4.962	PS428Q05	40	3.01	(0.311)	1.965
PF054Q01	20	3.72	(0.510)	2.282	PS438Q03	41	7.68	(0.624)	3.998
PF058Q01	20	3.14	(0.550)	2.458	PS465Q01	40	6.08	(0.495)	3.128
PF068Q01	20	3.07	(0.519)	2.321	PS514Q02	41	1.45	(0.241)	1.541
PF082Q01	20	3.90	(0.577)	2.580	PS514Q03	41	4.61	(0.408)	2.611
PF102Q02	20	7.21	(1.033)	4.621	PS519Q01	41	12.12	(0.972)	6.226
PF103Q01	20	3.38	(0.358)	1.603	PS519Q03	40	6.51	(0.773)	4.888
PF106Q01	20	3.42	(0.516)	2.309					

There were no such items in the computer-based mathematics and problem solving assessments. There was one item with $T_i > 7.5$ in the financial literacy and digital reading and there were two such items in the paper based domains of reading, mathematics and science. Most of the items in paper-based mathematics, computer-based mathematics and problem solving have a satisfactory $T_i < 3$ (highlighted in blue) which means that in these domains most of the items on average were coded consistently across all coders in all locales. Computer-based reading and paper-based domains of reading, science, and financial literacy were more difficult to code and as a result most of the items in these domains have $T_i > 3$.

Table 13.4 compares the international item reliability indices for link items between 2009 and 2012 cycles of PISA. It shows that the index is a stable measure. The change between cycles is statistically significant only for three items: the coding of the reading items PR220Q01 and PR432Q05 improved in 2012 and the coding of the mathematics item PM828Q01 become less consistent.

Table 13.4 Comparison of OECD/International item reliability indices (Ti) for link items between PISA 2009 and PISA 2012 cycles

Item	Ti_2012	S.E.	Ti_2009	S.E.	Z-value
Mathematics					
PM155Q01	1.17	(0.180)	1.61	(0.167)	-1.779
PM155Q02	3.50	(0.332)	4.03	(0.402)	-1.017
PM155Q03	5.65	(0.555)	5.18	(0.484)	0.639
PM406Q01	1.08	(0.165)	1.32	(0.123)	-1.170
PM406Q02	2.44	(0.333)	2.21	(0.255)	0.539
PM446Q02	1.24	(0.572)	0.84	(0.114)	0.697
PM462Q01	1.65	(0.213)	1.80	(0.172)	-0.538
PM828Q01	6.24	(0.475)	4.41	(0.377)	3.016
Reading					
PR220Q01	3.87	(0.350)	4.98	(0.428)	-2.001
PR404Q10A	4.13	(0.345)	4.75	(0.392)	-1.186
PR404Q10B	5.99	(0.494)	6.18	(0.525)	-0.254
PR406Q01	2.41	(0.266)	2.47	(0.209)	-0.163
PR406Q02	8.05	(0.725)	8.13	(0.699)	-0.080
PR406Q05	2.64	(0.321)	2.99	(0.296)	-0.797
PR412Q08	5.53	(0.385)	5.56	(0.500)	-0.045
PR420Q06	5.37	(0.473)	6.42	(0.570)	-1.411
PR432Q05	2.67	(0.220)	4.69	(0.487)	-3.784
PR437Q07	6.27	(0.490)	6.68	(0.583)	-0.532
PR446Q06	1.79	(0.298)	2.47	(0.312)	-1.590
PR453Q04	7.57	(0.666)	7.59	(0.626)	-0.028
PR453Q06	3.63	(0.380)	4.46	(0.382)	-1.527
PR455Q02	6.04	(0.504)	6.19	(0.498)	-0.214
PR456Q02	3.11	(0.400)	3.80	(0.356)	-1.286
PR456Q06	1.34	(0.140)	1.72	(0.182)	-1.651
PR466Q02	1.99	(0.226)	2.23	(0.227)	-0.747
Science					
PS131Q02	3.15	(0.373)	3.35	(0.334)	-0.399
PS131Q04D	3.77	(0.335)	4.12	(0.444)	-0.627
PS269Q01	2.15	(0.268)	2.22	(0.188)	-0.214
PS269Q03D	2.62	(0.323)	2.82	(0.368)	-0.416
PS326Q01	4.42	(0.408)	4.35	(0.413)	0.111
PS326Q02	4.18	(0.449)	3.77	(0.371)	0.713
PS408Q03	5.99	(0.516)	5.04	(0.515)	1.298
PS425Q03	6.95	(0.676)	7.22	(0.630)	-0.285
PS425Q04	3.31	(0.433)	3.51	(0.295)	-0.364
PS428Q05	3.01	(0.311)	3.61	(0.399)	-1.184
PS438Q03	7.68	(0.624)	6.88	(0.588)	0.928
PS465Q01	6.08	(0.495)	5.95	(0.546)	0.165
PS514Q02	1.45	(0.241)	1.40	(0.160)	0.169
PS514Q03	4.61	(0.408)	4.39	(0.402)	0.490
PS519Q01	12.12	(0.972)	12.06	(1.028)	0.063
PS519Q03	6.51	(0.773)	6.09	(0.600)	0.325

Let C be a set of σ test languages within the economy participating in the reliability exercise for the domain, D, and δ be the number of items in the domain D retained in the locale (see the list of all items deleted at the national level in Table 12.10, Chapter 12). The average disagreement for each participant across all items in each of the domains is then presented by national domain reliability indices N_{cD}.

13.5

$$N_{cD} = \sum_{c\in C,\delta\in D} \frac{S_{ic}}{\sigma\delta}$$

The national domain indices N_{cD} are presented in Table 13.5 for paper-based domains and in Table 13.6 for computer-based domains. $N_{cD} > 7.5$ are highlighted in grey as unusually high and $N_{cD} < 0.5$ are highlighted in blue as unusually low. These tables confirm the observation from the previous table that some domains were easier to code consistently than others. The most consistent were the mathematics domains (both paper-based and computer-based) and problem solving with average N_{cD} across all participants less than 3. Paper-based domains of reading, science and financial literacy were coded less consistently with average N_{cD} across all participants around 4.5 (for paper-based reading $N_{cD} = 4.37$, for paper-based science $N_{cD} = 4.66$ and for financial literacy $N_{cD} = 4.51$). The most difficult domain to code was digital reading with $N_{cD} = 6.03$. This was based on the existence of only four expert-coded items in the digital reading component and should be treated with caution. The online coding software provided a highly

Table 13.5 National domain reliability indices (N_{cD}) for paper-based domains

		Mathematics		Reading		Science		Financial literacy	
		N_{cD}	S.E.	N_{cD}	S.E.	N_{cD}	S.E.	N_{cD}	S.E.
OECD	Australia	2.39	(0.402)	5.22	(0.740)	5.79	(0.802)	5.75	(0.698)
	Austria	1.80	(0.433)	4.77	(0.543)	5.25	(0.881)		
	Belgium[1]	2.70	(0.399)	2.55	(0.311)	4.62	(0.576)	3.65	(0.673)
	Canada	4.43	(0.645)	5.99	(0.659)	10.30	(1.113)		
	Chile	3.34	(0.839)	6.03	(1.040)	8.47	(1.306)		
	Czech Republic	3.04	(0.757)	5.96	(1.109)	3.44	(0.455)	4.16	(0.816)
	Denmark	2.55	(0.414)	5.53	(0.807)	5.93	(0.942)		
	Estonia	2.71	(0.580)	3.46	(0.877)	4.89	(0.875)	5.94	(1.407)
	Finland	1.49	(0.336)	4.33	(0.654)	5.70	(0.865)		
	France	3.17	(0.764)	6.70	(0.909)	7.41	(1.476)	7.21	(1.167)
	Germany	3.24	(0.682)	5.86	(0.662)	5.60	(0.754)		
	Greece	1.07	(0.235)	1.68	(0.309)	2.81	(0.328)		
	Hungary	1.49	(0.357)	3.47	(0.554)	1.98	(0.449)		
	Iceland	2.26	(0.508)	4.87	(0.802)	4.16	(0.841)		
	Ireland	3.48	(0.742)	4.94	(0.878)	6.41	(0.969)		
	Israel	2.66	(0.345)	4.25	(0.654)	5.02	(0.664)	4.20	(0.549)
	Italy	2.48	(0.746)	3.17	(0.446)	6.09	(1.061)	2.56	(0.529)
	Japan	1.02	(0.240)	1.48	(0.235)	2.89	(0.398)		
	Korea	0.73	(0.211)	1.52	(0.310)	1.31	(0.287)		
	Luxembourg	1.84	(0.312)	2.60	(0.337)	3.79	(0.514)		
	Mexico	4.30	(1.430)	7.46	(1.187)	6.71	(1.066)		
	Netherlands	2.81	(0.462)	4.58	(0.801)	6.23	(1.091)		
	New Zealand	2.70	(0.495)	4.57	(0.678)	4.34	(0.635)	4.76	(0.691)
	Norway	3.17	(0.677)	3.81	(0.451)	6.98	(1.446)		
	Poland	2.74	(0.467)	2.41	(0.396)	2.81	(0.419)	2.56	(0.357)
	Portugal	1.41	(0.316)	4.92	(0.808)	1.39	(0.275)		
	Slovak Republic	1.12	(0.271)	5.07	(0.894)	3.55	(0.630)	1.30	(0.259)
	Slovenia	1.81	(0.436)	3.05	(0.400)	4.10	(0.750)	3.89	(0.827)
	Spain	2.48	(0.320)	5.87	(0.558)	5.53	(0.742)	4.02	(0.653)
	Sweden	3.57	(0.576)	4.26	(0.536)	4.85	(0.969)		
	Switzerland	1.39	(0.238)	2.23	(0.316)	2.51	(0.451)		
	Turkey	1.30	(0.301)	3.40	(0.752)	0.96	(0.237)		
	United Kingdom	2.01	(0.306)	5.17	(0.599)	5.34	(0.544)		
	United States	2.12	(0.447)	4.25	(0.683)	6.03	(0.814)	5.34	(1.189)
Partners	Albania	0.07	(0.070)	0.00	(0.000)	0.52	(0.354)		
	Argentina	0.23	(0.066)	0.19	(0.063)	0.59	(0.094)		
	Brazil	3.76	(1.160)	7.99	(1.096)	8.40	(1.439)		
	Bulgaria	3.73	(0.838)	6.92	(1.201)	5.78	(0.971)		
	Colombia	2.66	(0.640)	2.87	(0.385)	4.40	(0.664)	4.28	(0.953)
	Costa Rica	0.35	(0.150)	8.04	(1.458)	12.98	(2.240)		
	Croatia	1.67	(0.347)	5.59	(1.066)	6.52	(1.277)	6.44	(1.036)
	Cyprus[2, 3]	0.85	(0.184)	0.27	(0.077)	0.60	(0.102)		
	Hong Kong-China	2.61	(0.524)	4.28	(0.863)	7.70	(1.346)		
	Indonesia	4.86	(1.120)	17.47	(1.763)	11.57	(1.512)		
	Jordan	0.17	(0.051)	0.27	(0.090)	0.50	(0.120)		
	Kazakhstan	0.61	(0.090)	0.76	(0.118)	1.99	(1.018)		
	Latvia	4.04	(0.903)	11.75	(1.512)	10.45	(1.605)	9.29	(1.560)
	Lithuania	1.95	(0.435)	3.26	(0.512)	3.16	(0.591)		
	Macao-China	1.44	(0.260)	3.90	(0.524)	1.38	(0.169)		
	Malaysia	5.50	(0.897)	9.09	(0.932)	8.58	(0.974)		
	Montenegro	1.37	(0.373)	6.77	(1.172)	9.26	(1.182)		
	Peru	1.38	(0.673)	2.65	(0.393)	3.69	(0.715)		
	Qatar	0.67	(0.139)	1.56	(0.240)	0.48	(0.172)		
	Romania	0.32	(0.091)	6.31	(0.882)	0.75	(0.175)		
	Russian Federation	0.45	(0.155)	1.60	(0.309)	1.01	(0.180)	4.57	(0.737)
	Serbia	3.52	(0.750)	5.59	(0.769)	3.70	(0.502)		
	Shanghai-China	1.25	(0.375)	2.23	(0.482)	0.80	(0.177)	1.32	(0.448)
	Singapore	0.30	(0.084)	0.08	(0.046)	0.46	(0.191)		
	Chinese Taipei	1.28	(0.285)	2.15	(0.433)	2.99	(0.712)		
	Thailand	0.00	(0.000)	0.12	(0.056)	0.13	(0.073)		
	Tunisia	2.81	(0.737)	6.59	(1.022)	12.38	(1.875)		
	United Arab Emirates	1.93	(0.402)	2.60	(0.325)	3.25	(0.473)		
	Uruguay	2.44	(0.805)	6.19	(0.751)	3.32	(0.461)		
	Viet Nam	2.83	(1.091)	7.17	(2.461)	7.76	(1.689)		
	Mean (all participants)	**2.12**		**4.37**		**4.66**		**4.51**	
	SD (all participants)	**1.26**		**2.93**		**3.14**		**2.00**	

1. Only the Flemish community of Belgium participated in the financial literacy assessment.

2. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 13.6 **National domain reliability indices (N_{cD}) for computer-based domains**

		Problem solving		Mathematics		Digital reading	
	Participant	N_{cD}	S.E.	N_{cD}	S.E.	N_{cD}	S.E.
OECD	Australia	3.56	(1.00)	3.73	(1.25)	11.08	(1.08)
	Austria	2.73	(1.06)	2.06	(1.00)	5.58	(1.06)
	Belgium	2.07	(0.63)	3.92	(1.03)	5.76	(0.87)
	Canada	3.07	(0.61)	3.19	(0.70)	9.63	(2.30)
	Chile	3.63	(1.25)	4.17	(1.22)	8.08	(0.80)
	Czech Republic	2.60	(0.57)				
	Denmark	2.52	(0.66)	4.47	(1.87)	10.05	(2.21)
	Estonia	2.42	(0.60)	3.13	(0.96)	5.79	(1.26)
	Finland	1.03	(0.46)				
	France	4.51	(2.34)	5.47	(2.21)	13.20	(1.89)
	Germany	3.10	(0.88)	3.88	(1.82)	8.23	(0.96)
	Hungary	2.27	(0.71)	1.97	(0.56)	4.14	(0.67)
	Ireland	2.99	(0.97)	2.61	(0.48)	8.60	(1.38)
	Israel	2.30	(0.45)	4.00	(1.02)	6.09	(1.40)
	Italy	3.42	(0.87)	4.81	(1.73)	3.19	(0.57)
	Japan	2.14	(0.47)	2.26	(1.10)	4.04	(0.97)
	Korea	0.65	(0.36)	0.73	(0.18)	2.33	(0.79)
	Netherlands	3.57	(0.77)				
	Norway	2.67	(0.89)	3.62	(1.20)	4.40	(1.27)
	Poland	1.09	(0.27)	1.13	(0.92)	5.00	(0.94)
	Portugal	4.18	(1.31)	1.55	(0.31)	6.85	(2.59)
	Slovak Republic	3.77	(1.64)	1.97	(1.03)	6.37	(0.97)
	Slovenia	0.96	(0.37)	1.01	(0.20)	2.14	(0.61)
	Spain	1.33	(0.26)	1.62	(0.48)	5.26	(0.68)
	Sweden	2.68	(0.70)	6.08	(1.69)	6.51	(1.37)
	Turkey	3.05	(0.98)				
	United Kingdom[1]	5.00	(1.90)				
	United States	3.24	(1.04)	4.67	(1.82)	6.20	(1.13)
Partners	Brazil	2.41	(1.04)	1.99	(0.84)	5.32	(0.81)
	Bulgaria	4.40	(1.28)				
	Colombia	3.12	(0.85)	4.51	(2.11)	6.71	(0.56)
	Croatia	5.28	(2.51)				
	Cyprus[2, 3]	1.67	(0.48)				
	Hong Kong-China	2.55	(1.19)	4.50	(1.86)	5.43	(2.95)
	Macao-China	0.06	(0.06)	0.45	(0.30)	2.47	(0.62)
	Malaysia	2.49	(0.53)				
	Montenegro	7.08	(2.49)				
	Russian Federation	1.38	(0.20)	2.35	(0.93)	4.86	(0.94)
	Serbia	2.21	(0.88)				
	Shanghai-China	1.97	(0.46)	3.29	(1.42)	5.30	(0.78)
	Singapore	1.60	(0.80)	2.71	(1.64)	6.48	(3.16)
	Chinese Taipei	2.15	(1.17)	1.56	(0.65)	4.08	(0.67)
	United Arab Emirates	1.68	(0.44)	2.11	(0.50)	3.91	(0.68)
	Uruguay	3.27	(0.73)				
	Mean	**2.72**		**2.99**		**6.03**	
	SD	**1.31**		**1.44**		**2.52**	

1. England only.

2. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

3. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

sophisticated means of coding student responses, which accommodate all but four of the reading items, and these were the items requiring the most complex judgements. Historically, reading items have always been more difficult to code than mathematics items.

INTERNATIONAL CODER REVIEW

Control scripts

With the introduction of the online coding system the opportunity was provided in the PISA 2012 administration to develop an objective alternative for the international coder review. The item developers provided responses for constructed response expert items for each domain and correct coding for each response. These responses are referred to as *control scripts* in this chapter. National Centres translated control scripts and scanned translations into the online coding system where they were presented to coders as student responses, indistinguishable from other student responses.

This was done for all domains that were coded on line (all computer-based domains and for some countries for paper-based domains). Control scripts were provided to allow for international bias analysis by comparison of codes given to the same response by coders from different National Centres on the one hand and by item developers on the other hand. Table 13.7 shows participation in the control-script exercise by domain. Fifty-five locales coded control scripts for at least one domain. The use of control scripts enabled National Centres to monitor the quality of their coding in real time, since the online coding system allowed National Centres to re-train coders when observed discrepancies between coders and provided control scripts were high.

Table 13.7 Participation in control scripts bias analysis by domain

		Paper-based domains		Computer-based domains	
	Locale	Mathematics, reading, science	Financial literacy	Mathematics, digital reading	Problem solving
OECD	Australia-English	Y	Y	Y	Y
	Austria-German	Y		Y	Y
	Belgium-Flemish			Y	Y
	Belgium-French			Y	Y
	Canada-English			Y	Y
	Canada-French			Y	Y
	Chile-Spanish			Y	Y
	Czech Republic-Czech				Y
	Denmark-Danish	Y		Y	Y
	Estonia-Estonian			Y	Y
	Estonia-Russian			Y	Y
	Finland-Finnish				Y
	France-French			Y	Y
	Germany-German			Y	Y
	Hungary-Hungarian			Y	Y
	Iceland-Icelandic	Y			
	Ireland-English			Y	Y
	Israel-Arabic	Y	Y	Y	Y
	Israel-Hebrew	Y	Y	Y	Y
	Italy-Italian			Y	Y
	Japan-Japanese			Y	Y
	Korea-Korean	Y		Y	Y
	Netherlands-Dutch				Y
	Norway-Norwegian			Y	Y
	Poland-Polish			Y	Y
	Portugal-Portuguese			Y	Y
	Slovak Republic-Slovak			Y	Y
	Slovenia-Slovenian			Y	Y
	Spain-Basque			Y	Y
	Spain-Catalan			Y	Y
	Spain-Spanish			Y	Y
	Sweden-Swedish	Y		Y	Y
	Switzerland-French	Y			
	Switzerland-German	Y			
	Turkey-Turkish				Y
	United Kingdom-English				Y[3]
	United States-English	Y	Y	Y	Y
Partners	Brazil-Portuguese			Y	Y
	Bulgaria-Bulgarian				Y
	Colombia-Spanish	Y	Y	Y	Y
	Croatia-Croatian				Y
	Cyprus-English[1, 2]				Y
	Cyprus-Greek[1, 2]				Y
	Hong Kong-China-Chinese			Y	Y
	Macao-China-Chinese			Y	Y
	Malaysia-English				Y
	Malaysia-Malay				Y
	Montenegro-Montenegrin				Y
	Russian Federation-Russian			Y	Y
	Shanghai-China-Chinese			Y	Y
	Singapore-English			Y	Y
	Chinese Taipei-Chinese	Y		Y	Y
	United Arab Emirates-Arabic	Y		Y	Y
	United Arab Emirates-English	Y		Y	Y
	Uruguay-Spanish	Y			Y

1. Footnote by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

3. England only.

The items and the number of control scripts for each domain are listed in Table 13.8. The number of control scripts per item was determined by item developers. To avoid dependency between scripts each script represented a different type of answer. This approach, however, often provided only single digit number of scripts per item, which is essentially equivalent to the single digit number of student responses per item per locale. Because the number of scripts available was relatively small, and the number of countries participating in this new approach to international coder review was limited, the volume of material generated through the use of control scripts was not sufficient to perform a robust analysis of the outcomes of the procedures used. Nevertheless, it is expected on the basis of the experience gained that higher levels of participation in future would lead to better data volumes and this would permit analysis of outcomes to be carried out. In future control scripts can be used if there are more constructed response items in each of the computer-based domains or if all participants in the paper-based domains use online coding or both. For the Field Trial, control scripts still can be used as an effective tool to improve coding guides and to identify items that are difficult to code.

Comparison of student achievement in constructed response and all other items

Since the use of control scripts did not provide data of sufficient volume for identification of bias (Table 13.8) a different statistical procedure was employed. In summary, the procedure compared two differences between student achievements in each of the 100 achievement categories. The difference I_j (j=1,…100) between achievement in constructed response and all other items internationally was used as a benchmark. This statistic was based on the plausible values for all PISA students who participated in the domain. It was compared to the differences L_{kj} between student achievement in constructed response and all other items in each participant *k*. This statistics was based on the plausible values for the

Table 13.8 **The list of items for which control scripts were provided**

Item ID	Number of control scripts	Item ID	Number of control scripts
Computer-based mathematics		**Paper-based mathematics**	
CM015Q03	10	PM00FQ01	8
CM028Q03	8	PM00KQ02	8
CM038Q05	8	PM155Q01	3
CM038Q06	9	PM155Q02	5
Problem solving		PM155Q03	4
CP002Q06	14	PM406Q01	4
CP018Q05	8	PM406Q02	4
CP034Q05	5	PM462Q01	6
CP036Q02	5	PM828Q01	2
CP036Q03	6	PM903Q01	8
CP041Q02	11	PM905Q02	8
Digital reading		PM906Q02	7
CR002Q05	16	PM949Q03	8
CR013Q07	14	PM953Q04	8
CR014Q01	17	PM955Q03	7
CR017Q07	18	PM961Q05	7
CR021Q08	19	PM991Q02	7
Science		**Paper-based reading**	
PS131Q02	9	PR404Q10A	5
PS131Q04	9	PR404Q10B	4
PS269Q01	10	PR406Q01	6
PS269Q03	9	PR406Q02	7
PS326Q01	8	PR406Q05	7
PS326Q02	7	PR412Q08	3
PS408Q03	8	PR420Q06	5
PS425Q03	8	PR420Q10	4
PS425Q04	9	PR432Q05	4
PS428Q05	9	PR437Q07	5
PS438Q03	10	PR446Q06	2
PS465Q01	10	PR453Q04	3
PS498Q04	10	PR453Q06	4
PS514Q02	10	PR455Q02	5
PS514Q03	10	PR456Q02	4
PS519Q01	10	PR456Q06	3
PS519Q03	8	PR466Q02	4
Financial literacy			
PF004Q03	12		
PF024Q02	13		
PF028Q02	13		
PF036Q01	19		
PF051Q01	15		
PF082Q01	14		
PF102Q02	17		

subset of students from this participant. The assumption was that if L_{kj} behaves statistically differently from $_{lj}$ persistently across j, it may partially be attributed to coding bias.

We know from previous research (Routitsky and Turner, 2003) that there can be differences in performance on items of different format (e.g. multiple choice and constructed response items) and that the magnitude of this difference varies for students of different abilities. Therefore, L_{kj} were expected to vary across achievement categories (j=1,…,100) within participants (as well as between them) and were compared to the corresponding I_j which was used as a benchmark.

In detail, the procedure was implemented as follows.

International item parameters were used for all domains. For paper-based mathematics, only common items were used; the items that were unique to standard booklets and items that were unique to easier booklets were excluded to facilitate comparison between countries that used easier booklets and countries that used standard booklets.

For each domain, the items in the item parameter file were divided into two groups. One group contained constructed response items ("CR" item group), and the other group contained the rest of the items ("Rest" item group). Item parameters of each group were adjusted to a parameter mean of zero nationally: if an item was deleted from participant data, a separate item parameter file was created by excluding this item and re-adjusting all item parameters to the mean of zero.

The ACER *ConQuest* (Adams, Wu and Wilson, 2012) programme file was created to estimate plausible values for student achievement based on each item group of each domain within each participant using a 2-dimensional model. For each domain the plausible values estimated by ACER *ConQuest* were read into SPSS© (2010) and processed as described below.

Let W_D be the weighted number of students for the domain D across all participants.

Let $\left\{RP_{s,i}\right\}$ ($s=1,\dots,W_D$; $i=1,\dots,5$), be a set of plausible values derived for the "Rest" item group of the domain D.

For each $i=1,\dots,5$ RP_{si} was sorted in ascending order and divided into 100 equally weighted sets A_{ji} ($j=1,\dots,100$; $i=1,\dots,5$) of the W_D=W_D/100 size. For each $i=1,\dots,5$ the new variable S_i was constructed. All students from A_{ji} were assigned $S_i = j$, meaning that according to the plausible value i the student belongs to the achievement group j. Note that for the same student the value of j could be different for different plausible values.

Let $\left\{CP_{s,i}\right\}$ ($s=1,\dots,W_D$; $i=1,\dots,5$), be a set of plausible values derived for the "CR" (constructed response) item group of the domain D.

For each set A_{ji} ($j=1,\dots,100$; $i=1,\dots,5$) the mean difference was calculated as follows

13.6

$$MI_{ji} = \frac{\sum_{s\in A_{ji}} v_s \left(CP_{s,i} - RP_{s,i}\right)}{w_D},$$

where v_s is a total student weight for students (see Chapter 8 for details about weight estimation).

The difference I_j (j=1,…,100) between achievement in constructed response and all other items internationally was calculated as the average between 5 differences MI_{ji}:

13.7

$$I_j = \frac{\sum_{i=1}^{5} MI_{ji}}{5}$$

I_j can be interpreted as achievement in constructed response items relative to the achievement in all other items and will be called in the rest of this chapter *relative international achievement*.

The differences L_{kj} between student achievement in constructed response and all other items in each participant k were calculated as follows.

Let $B_{k\,ji}$ be a subset of A_{ji} from a participant k: $B_{k\,ji} \subset A_{ji}$ and m_{kiD} the weighted number of students in this set. Then

13.8

$$ML_{kji} = \frac{\sum_{s \in B_{kji}} v_s \left(CP_{s,i} - RP_{s,i}\right)}{m_{kiD}}$$

13.9

$$L_{kj} = \frac{\sum_{i=1}^{5} ML_{kji}}{5}$$

L_{kj} can be interpreted as achievement in constructed response items relative to the achievement in all other items within the locale and will be called in the rest of this chapter *relative locale achievement*.

Standard errors for I_j and L_{kj} were calculated using the balanced repeated replication method. Standard errors were used to run z-tests with α=0.05 to find whether L_{kj} is significantly different from I_j. Z-test showed that the difference L_{kj} - I_j was statistically significantly different from zero for some j within some participant k. However, the differences were not systematic across different achievement groups j. Therefore, the next step was to identify the size of this difference for each participating country and economy. To identify the size of the difference between L_{kj} - I_j within a particular participant the following approach was employed.

Let CIL_{kj} be a lower boundary of the confidence interval of the difference L_{kj} - I_j. Then,

if $CIL_{kj} > 0$, the adjusted plausible values for constructed response items $RCP_{s,i}$ were computed for all plausible values $s \in B_{k\,ji}$ as

13.10

$$RCP_{s,i} = CP_{s,i} - CIL_{kj}$$

Let CIU_{kj} be an upper boundary of the confidence interval of the difference L_{kj} - I_j. Then,

if $CIL_{kj} < 0$, corrected plausible values for constructed response items $RCP_{s,i}$ were computed for all plausible values $s \in B_{k\,ji}$ as

13.11

$$RCP_{s,i} = CP_{s,i} - CIU_{kj}$$

Finally, if $CIL_{kj} < 0 < CIU_{kj}$,

13.12

$$RCP_{s,i} = CP_{s,i}$$

The adjusted plausible values for constructed response items $RCP_{s,i}$ were then compared to the initial plausible values $CP_{s,i}$ within each participating country/economy by calculating the average difference G_{kj} [13.14] and its standard error as well as standard deviation SD(G_{kj}) using the balanced repeated replication method.

13.13

$$MG_{kji} = \frac{\sum_{s \in B_{kji}} v_s \left(RCP_{s,i} - CP_{s,i}\right)}{m_{kiD}}$$

13.14

$$G_{kj} = \frac{\sum_{i=1}^{5} MG_{kji}}{5}$$

Table 13.9 Percent of students in the lowest level of proficiency and amount of difference between national and international relative achievement in constructed response items (G_{kj}) by domain

		Paper-based domains								Computer-based domains			
		Mathematics		Reading		Science		Financial literacy		Problem solving		Digital reading	
	Participant	% below Level 1	G_{kj}	% below Level 1a	G_{kj}	% below Level 1	G_{kj}	% below Level 1	G_{kj}	% below Level 1	G_{kj}	% below Level 2	G_{kj}
OECD	Australia	6.1	-0.13	4.01	-0.08	3.4	-0.14	3.39	-0.01	5.03	-0.02	12.46	-0.01
	Austria	5.7	-0.07	5.66	-0.03	3.6	-0.06			6.49	0.00	20.23	0.01
	Belgium[1]	7.0	-0.10	5.74	-0.05	5.8	-0.05	2.72	-0.03	9.08	0.00	17.19	0.00
	Canada	3.6	-0.16	2.86	-0.07	2.4	-0.14			5.10	-0.02	8.46	-0.01
	Chile	22.0	0.06	9.08	0.03	8.1	0.01			15.15	0.03	29.30	0.03
	Czech Republic	6.8	-0.06	4.12	-0.04	3.3	-0.16	3.09	-0.01	6.53	-0.02		
	Denmark	4.4	-0.11	3.90	-0.03	4.7	-0.05			7.30	0.00	14.23	0.01
	Estonia	2.0	-0.13	1.46	-0.08	0.5	-0.24	0.79	-0.03	4.01	-0.01	11.43	0.00
	Finland	3.3	-0.14	3.12	-0.09	1.8	-0.20			4.46	0.00		
	France	8.7	-0.06	6.99	-0.04	6.1	-0.06	8.68	0.00	6.63	-0.01	13.77	0.00
	Germany	5.5	-0.11	3.79	-0.07	2.9	-0.17			7.48	0.00	19.14	0.00
	Greece	14.5	0.01	8.47	-0.02	7.4	-0.01						
	Hungary	9.9	-0.03	5.95	-0.02	4.1	-0.06			17.22	0.01	32.48	0.03
	Iceland	7.5	-0.04	7.66	-0.02	8.0	-0.01						
	Ireland	4.8	-0.09	2.12	-0.10	2.6	-0.15			7.02	0.00	9.41	-0.01
	Israel	15.9	0.00	10.69	-0.03	11.2	-0.01	11.65	0.03	21.86	0.05	31.03	0.09
	Italy	8.5	-0.06	6.77	-0.07	4.9	-0.15	7.93	0.01	5.18	0.00	15.68	0.00
	Japan	**3.2**	**-0.23**	**3.06**	**-0.17**	**2.0**	**-0.32**			1.79	-0.07	4.92	-0.04
	Korea	**2.7**	-0.26	**2.15**	-0.15	**1.2**	-0.27			2.14	-0.05	3.95	-0.03
	Luxembourg	8.8	-0.03	8.33	-0.02	7.2	-0.03						
	Mexico	22.8	0.14	13.58	0.11	12.6	0.24						
	Netherlands	3.8	-0.11	3.72	-0.04	3.1	-0.11			7.36	0.00		
	New Zealand	7.5	-0.07	5.29	-0.05	4.7	-0.10	7.26	-0.01				
	Norway	7.2	-0.05	5.41	-0.05	6.0	-0.05			8.12	0.00	16.65	0.00
	Poland	3.3	-0.10	2.47	-0.09	1.3	-0.16	1.88	-0.01	10.04	0.00	22.39	0.00
	Portugal	8.9	-0.05	6.47	-0.02	4.7	-0.04			6.48	0.00	19.16	0.01
	Slovak Republic	11.1	-0.02	12.00	0.00	9.2	0.00	10.75	0.01	10.72	0.01	22.56	0.01
	Slovenia	5.1	-0.07	6.17	0.00	2.4	-0.03	5.32	0.00	11.39	0.02	25.12	0.04
	Spain	7.8	-0.05	5.75	-0.06	3.7	-0.12	4.94	0.00	13.14	0.00	26.16	0.01
	Sweden	9.5	-0.03	8.84	-0.03	7.3	-0.04			8.82	0.00	16.72	0.00
	Switzerland	3.6	-0.13	3.43	-0.06	3.0	-0.11						
	Turkey	15.5	0.00	5.06	-0.01	4.4	0.01			10.98	0.06		
	United Kingdom[2]	7.8	-0.06	5.44	-0.06	4.3	-0.11			5.55	-0.01		
	United States	8.0	-0.04	4.32	-0.02	4.2	-0.02	6.03	0.00	5.66	-0.01	12.61	0.00
Partners	Albania	32.5	0.12	27.98	0.11	23.5	0.16						
	Argentina	34.9	0.14	25.86	0.11	19.8	0.15						
	Brazil	35.2	0.20	18.77	0.16	18.6	0.28			21.89	0.06	37.16	0.09
	Bulgaria	20.0	0.02	20.80	0.01	14.4	0.01			33.33	0.24		
	Colombia	41.6	0.24	20.41	0.09	19.8	0.18	32.64	0.09	33.16	0.12	54.85	0.12
	Costa Rica	23.6	0.11	8.12	0.04	8.6	0.11						
	Croatia	9.5	-0.02	4.75	-0.03	3.2	-0.05	5.28	0.00	12.05	0.01		
	Cyprus[3, 4]	19.0	0.02	15.81	0.09	14.4	0.05			19.55	0.11		
	Hong Kong-China	**2.6**	**-0.29**	**1.51**	**-0.22**	**1.2**	**-0.37**			3.33	-0.03	7.57	-0.03
	Indonesia	42.3	0.22	20.41	0.16	24.7	0.36						
	Jordan	36.5	0.20	22.40	0.09	18.2	0.15						
	Kazakhstan	14.5	0.05	21.50	0.20	11.3	0.15						
	Latvia	4.8	-0.05	4.38	-0.04	1.8	-0.07	1.97	-0.02				
	Lithuania	8.7	-0.03	5.55	-0.01	3.4	-0.05						
	Macao-China	**3.2**	**-0.17**	2.46	-0.08	1.4	-0.17			1.55	-0.04	6.96	0.00
	Malaysia	23.0	0.08	22.23	0.10	14.5	0.12			22.66	0.14		
	Montenegro	27.5	0.08	17.55	0.05	18.7	0.14			30.00	0.17		
	Peru	47.0	0.27	30.43	0.18	31.5	0.39						
	Qatar	47.0	0.26	32.55	0.23	34.6	0.36						
	Romania	14.0	0.02	12.88	0.02	8.7	0.05						
	Russian Federation	7.5	-0.05	6.29	0.00	3.6	-0.03	5.53	0.00	6.76	0.00	24.61	0.03
	Serbia	15.5	0.01	11.86	0.02	10.3	0.03			10.27	0.03		
	Shanghai-China	**0.8**	**-0.52**	**0.39**	**-0.32**	**0.3**	**-0.52**	**0.32**	**-0.29**	3.09	-0.01	7.88	-0.01
	Singapore	**2.2**	**-0.34**	**2.41**	-0.17	**2.2**	-0.27			2.01	-0.05	4.35	-0.08
	Chinese Taipei	**4.5**	**-0.27**	3.05	-0.11	1.6	-0.19			3.44	-0.01	11.08	0.00
	Thailand	19.1	0.05	8.87	0.05	7.0	0.05						
	Tunisia	36.5	0.17	21.68	0.07	21.3	0.19						
	United Arab Emirates	20.5	0.06	13.71	0.06	11.3	0.07			30.28	0.17	50.48	0.20
	Uruguay	29.2	0.09	21.13	0.06	19.7	0.10			32.39	0.22		
	Viet Nam	3.6	-0.13	1.60	-0.11	0.9	-0.22						
	R^2	**0.77**		**0.74**		**0.75**		**0.29**		**0.87**		**0.82**	
	R^2 adjusted for outliers	**0.94**		**0.82**		**0.83**		**0.91**		**N/A**		**N/A**	

1. Only the Flemish community of Belgium took part in the assessment of financial literacy.

2. Only England took part in the assessment of problem solving.

3. Footnote by Turkey: The information in this document with reference to « Cyprus » relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

4. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

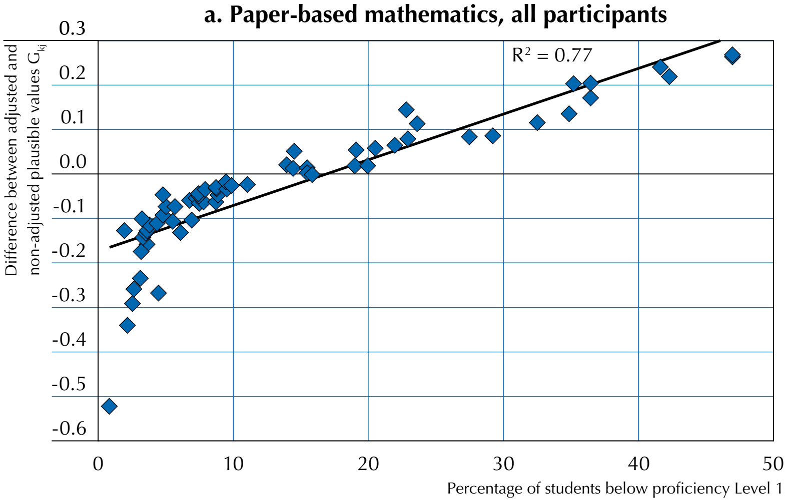
Due to the way adjusted plausible values were calculated, G_{kj} can be interpreted as the difference between relative national and international constructed response achievement (that is achievement in constructed response items relative to the achievement in all other items). Index G_{kj} was calculated for all domains except computer-based mathematics because computer-based mathematics had only 4 constructed response items; and for all participants except Liechtenstein. Data from only four items were deemed to be insufficient to calculate plausible values for all students. The number of students in data for Liechtenstein (293) was insufficient to estimate results separately for each type of item in each of 100 sets B_{kji}.

As mentioned earlier, we know from previous research that differential behaviour of various item formats depends on the level of a student's achievement (Routitsky and Turner, 2003). Thus, we would like to see how much of the variation in difference between national and international relative constructed response achievement G_{kj} can be explained by the percent of students in the lowest level of proficiency for each domain before we ascribe responsibility for any of this variation to country specific coding bias. The levels of proficiency are described in Volume I of the *PISA 2012 Results* (OECD, 2014). The lowest level of proficiency was chosen because students at this level are most likely to skip constructed response items and so would be least affected by coding bias and, therefore, correlation between the percentage of students in the lowest level of achievement and G_{kj} will be least confounded by coding bias. Table 13.9 shows side-by-side for each participant (except Liechtenstein) the percentage of students in the lowest level of proficiency for each domain except computer-based mathematics.

Figure 13.1a illustrates the relationship between G_{kj} and the percentage of students below proficiency level 1 for paper-based mathematics. It shows that 77% in G_{kj} variation is explained by the proportion of low achieving students in the country. G_{kj} shows that students from low achieving countries are achieving relatively better on the constructed response items than the students from the high achieving countries (relative to their achievement on all other items). One possible explanation of this could be that low achieving countries have some positive bias towards their students or high achieving countries have some negative bias towards their students or both. Eliminating this, were it the case, may only increase the distance between countries not the general ranking. However an alternative explanation is that G_{kj} is higher in the low achieving countries due to the fact that their achievement in all other items is so much lower.

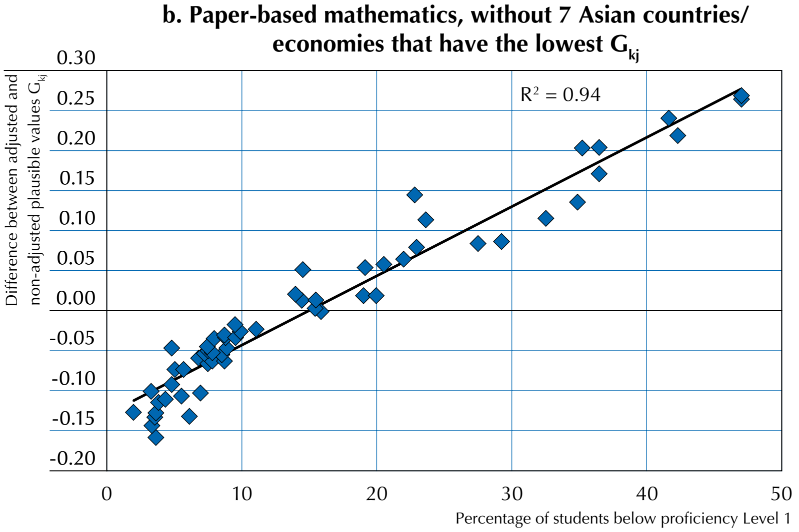
■ Figure 13.1 [Part 1/2] ■

Relationship between G_{kj} and percentage of students below proficiency Level 1 for paper-based mathematics



■ Figure 13.1 [Part 2/2] ■

Relationship between G_{kj} and percentage of students below proficiency Level 1 for paper-based mathematics



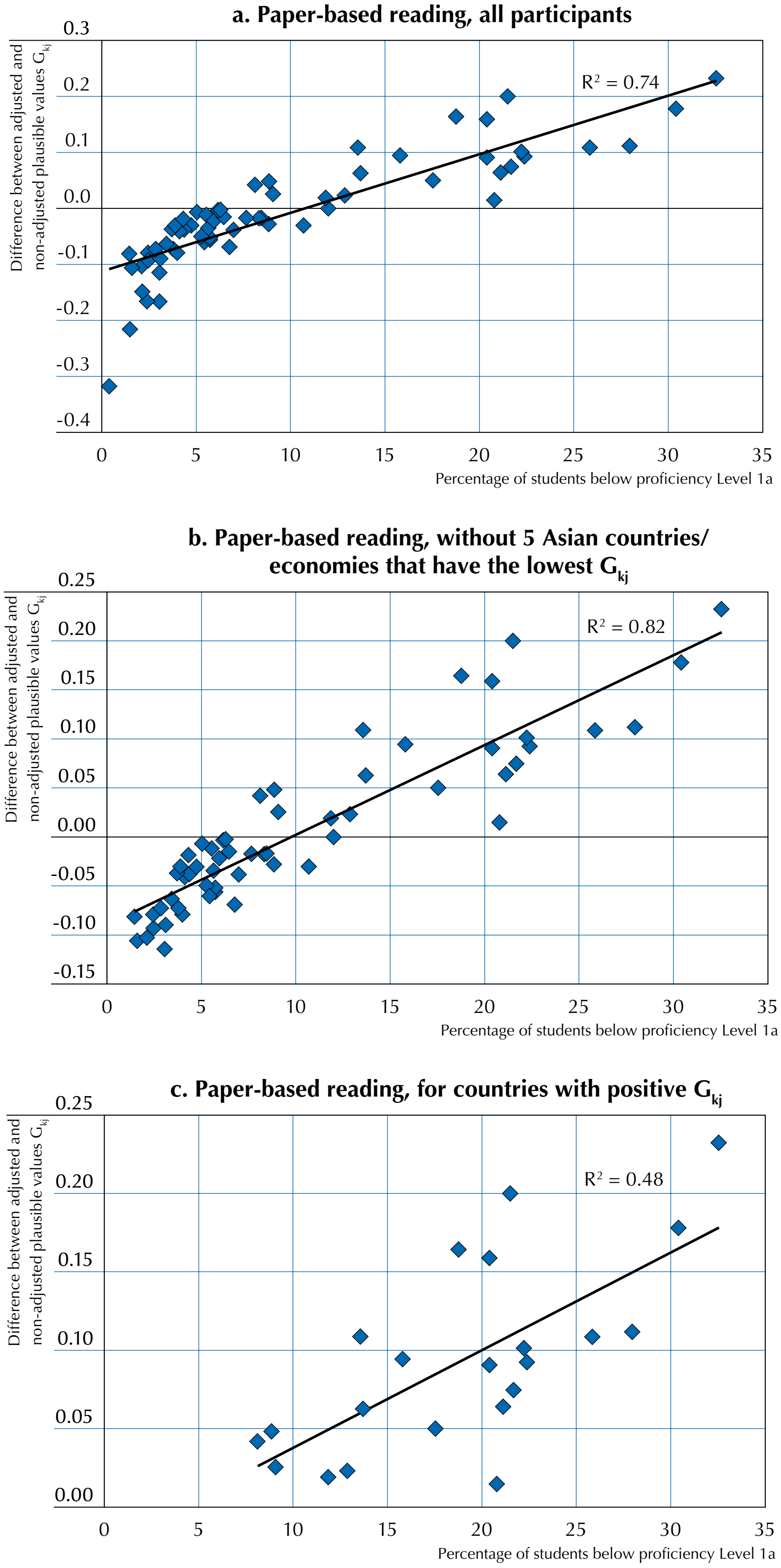
b. Paper-based mathematics, without 7 Asian countries/ economies that have the lowest G_{kj}

This is indeed the case and this would explain why the correlation with the proportion of students from the lowest level of proficiency is so high. We can also see some outliers in the left bottom part of the graph. These outliers belong to seven Asian participants that have G_{kj} ranging from -0.52 to -0.17 and a percentage of students below proficiency Level 1 ranging from 0.8% to 4.5% (Figure 13.1a and Table 13.9). These participants are Shanghai-China, Singapore, Hong Kong-China, Chinese Taipei, Korea, Japan and Macao-China, and are highlighted in bold in Table 13.9. Numerous researches comparing education in eastern and western countries (Leung et al., 2006) noticed that curriculum, teaching methods and assessment practices in these participants are different from those in other regions and have some similarities with each other. One possibility is that it is these factors that contribute to the variation in G_{kj} above and beyond the variation explained by the percentage of students in the lower level of achievement. The mechanism for this, however, is unclear and other reasons should be explored in the future. If we calculate R^2 without the above seven Asian participants, we can see that for the rest of PISA participants the proportion of low achieving students explains 94 % of variation in G_{kj} (Figure 13.1b).

There are similar results for paper-based reading and science. Figure 13.2 shows that 74 % in G_{kj} variation for reading is explained by the proportion of low achieving students in the country and Figure 13.3 shows that 75% in G_{kj} variation for science is explained by the proportion of low achieving students in the country.

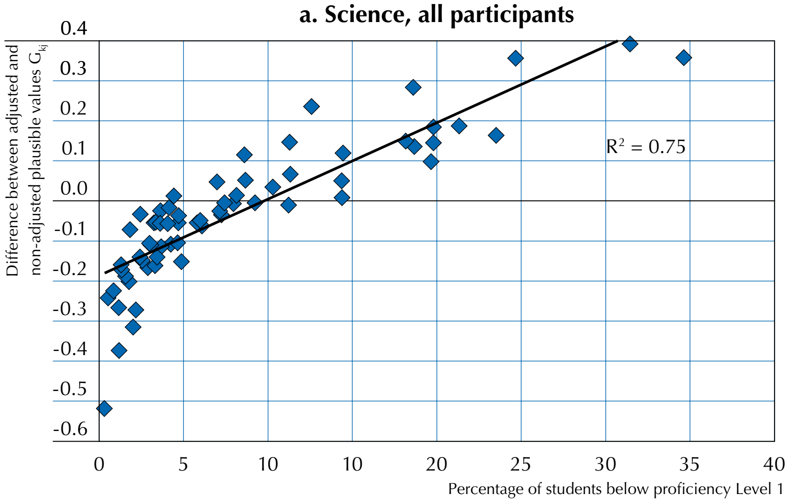
■ Figure 13.2 ■

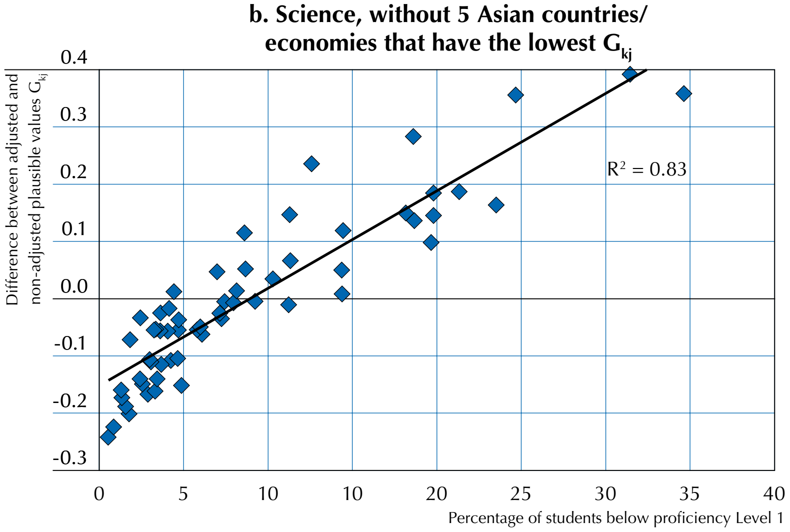
Relationship between G_{kj} and percentage of students below proficiency Level 1a for paper-based reading



■ Figure 13.3 ■

Relationship between G_{kj} and percentage of students below proficiency Level 1 for science





Results for financial literacy (Figure 13.4) seemed to be different but if Shanghai-China – which is the only and very clear outlier – is not taken into account, for the rest of participating countries R^2=91%, which is comparable to the paper-based mathematics result.

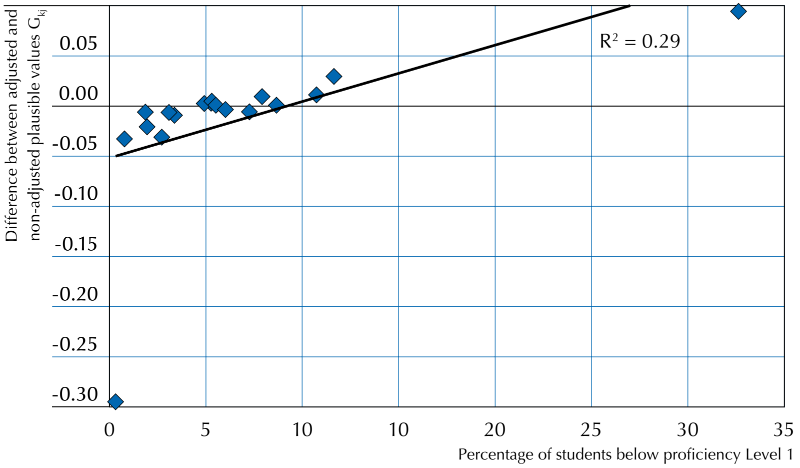
For the computer-based domains of problem solving and digital reading (Figures 13.5 and 13.6) there are no clear outliers and R^2 is higher than non-adjusted R^2 for paper-based domains. For problem solving 87% in G_{kj} variation is explained by the proportion of low achieving students in the country and for digital reading 82% in G_{kj} variation is explained by the proportion of low achieving students in the country.

Given that in addition to the differences between the percentage of students in different proficiency levels, there are some curriculum, teaching methods and assessment practices differences between PISA participants that can contribute to the variation in G_{kj} beyond and above the variation that is attributed to the percent of students in the lowest proficiency level, we can’t conclude that there is a bias in coding of constructed response items in any particular PISA economy.

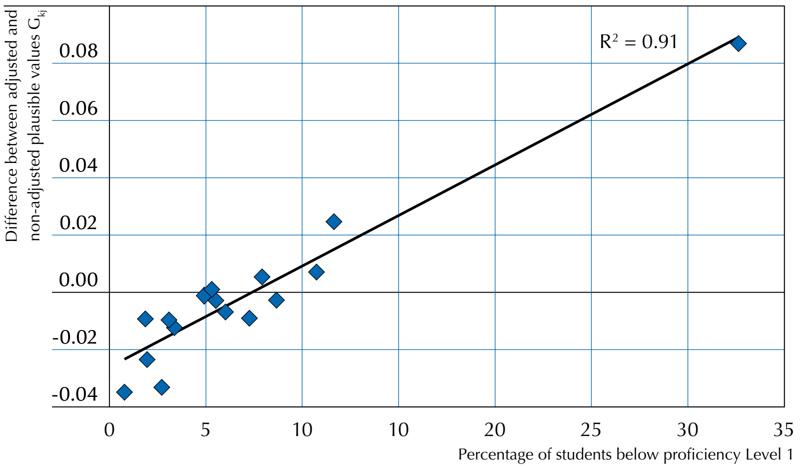
■ Figure 13.4 ■

Relationship between G_{kj} and percentage of students below proficiency Level 1 for financial literacy

a. Financial literacy, all participants



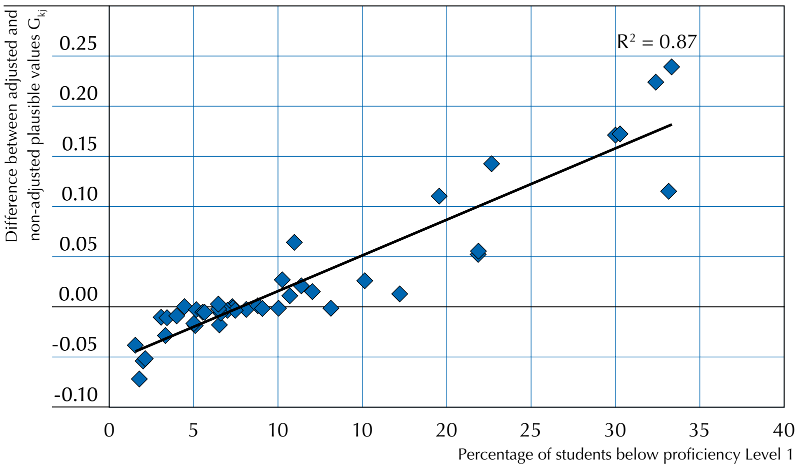
b. Financial literacy, excluding Shanghai-China



■ Figure 13.5 ■

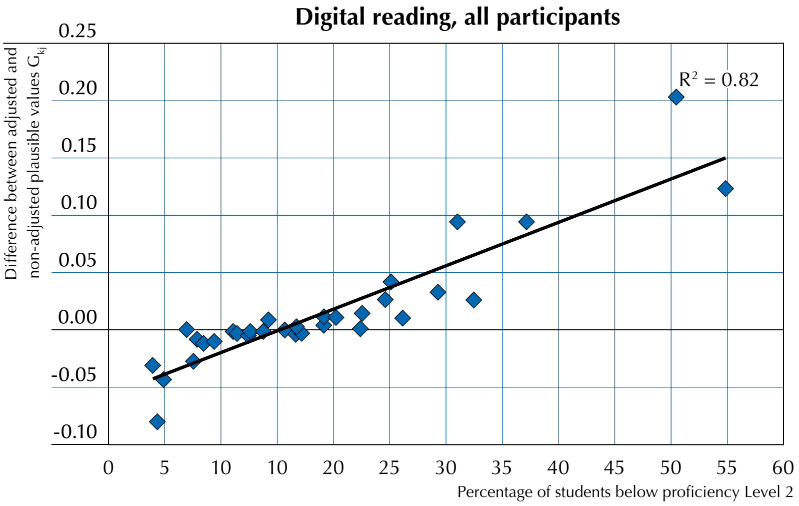
Relationship between G_{kj} and percent of students below proficiency Level 1 for problem solving

Problem solving, all participants



■ Figure 13.6 ■

Relationship between G_{kj} and percentage of students below proficiency Level 2 for digital reading



Note

1. Some items have been removed from analysis from some locales during adjudication process due to printing, translation and other errors (see Table 12.10, in Chapter 12, for the complete list of such items).

References

Adams, R., M. Wu, and **M. Wilson** (2012), *ACER ConQuest 3.1*, ACER, Melbourne.

Leung, F. K. S., K.D. Graf and **F. J. Lopez-Real** (ed.) (2006), "Mathematics Education in Different Cultural Traditions: A Comparative Study of East Asia and the West: The 13th ICMI Study", New ICMI Study Series, Volume 9, Springer, New York.

OECD (2014), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

Routitsky, A. and **R. Turner** (2003), "Item Format Types and their Influence on Cross-national Comparisons of Student Performance", Presentation given to the Annual Meeting of the American Educational Research Association (AERA) in Chicago, USA. Retrieved from *http://works.bepress.com/cgi/viewcontent.cgi?article=1013&context=alla_routitsky*

SPSS, IBM (2010), SPSS for Windows® (version 19), SPSS. Inc., Chicago, Illinois.

14

Data Adjudication

INTRODUCTION

The PISA Technical Standards (see Annex F) specify the way in which PISA must be implemented in each country. The international contractor has to monitor the implementation in each country and adjudicate on the countries' adherence to the standards. This chapter describes the process used to adjudicate the implementation of PISA 2012 in each of the entities (i.e. the participating countries, economies and adjudicated regions[1]) and it gives the outcomes of data adjudication that are mainly based on the following aspects:

- the extent to which each adjudicated entity met PISA sampling standards;
- the outcomes of the adaptation, translation and verification process;
- the outcomes of the National Centre and PISA Quality Monitoring visits and interviews;
- the quality and completeness of the submitted data; and
- the outcomes of the international coding review.

The areas covered in the PISA 2012 Technical Standards include the following:

Data Standards

- Target population and sampling
- Language of testing
- Field Trial participation
- Adaptation of tests, questionnaires and manuals
- Translation of tests, questionnaires and manuals
- Test administration
- Implementation of national options
- Security of the material
- Quality monitoring
- Printing of material
- Response coding
- Data submission

Management standards

- Communication with the international contractors
- Notification of international and national options
- Schedule for submission of materials
- Drawing samples
- Management of data
- Archiving of materials

National involvement standards

- National feedback

Implementing the standards – quality assurance

National Project Managers of participating countries, economies and adjudicated regions are responsible for implementing the standards based on the international contractor's advice as contained in the various operational manuals and guidelines. Throughout the cycle of activities for each PISA survey the international contractor 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 international contractor staff the opportunity to ensure that PISA implementation was planned in accordance with the PISA 2012 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 international contractor monitors a country's implementation of the data collection procedures from a range of perspectives and from processes occurring during many stages of the PISA cycle. These perspectives include monitoring a country's adherence to the deadlines, communication from the sampling contractor about each country's sampling plan, information from the language verification team, data from the PISA Quality Monitors and National Centre Quality Monitors, and information gleaned from direct interviews at National Project Manager and Marker Training meetings. The information is combined together in the database so that:

- indications of non-compliance with the standards can be identified early on in order to enable rectifying measures;
- the point at which the problem occurred can be easily identified; and
- information relating to the same PISA standard can be cross-checked between different areas or sources.

Many of these data collection procedures refer to specific milestone documents, specified in the National Project Manager's Manual and the Sampling Manual in particular. These are procedures that the international contractor requires for Field Trial and Main Survey preparation from each National Centre. The data adjudication process provides a motivation for collating and summarising the specific information relating to PISA standards collected in these documents, combined with information collected from specific quality monitoring procedures such as the national centre quality monitoring interview, PISA quality monitor visits and from information in the submitted data.

The quality monitoring information was collected from the following main administrative areas covering various quality monitoring instruments:

- International contractor administration and management: information relating to administration processes, agreement of adaptation spreadsheets, submission of information.
- Data analysis: information from the dodgy item reports, Field Trial sample, item information for cleaning.
- Field operations – manuals: information from the agreement of adaptations to test administration procedures and field operations.
- Final Optical Check team: information from the pre- and post-Main Survey Final Optical Checks of Main Survey booklets.
- Main Survey review: information provided by the National Project Managers in the Main Survey review process.
- National Centre quality monitoring: information gathered during the pre-Main Survey National Centre Quality Monitoring visits, and through interviews conducted during meetings of National Project Managers or at other times.
- PISA Quality Monitor country reports: information gathered via the test session reports from PISA Quality Monitors and through their interviews with School Co-ordinators.
- Sampling: information from the submitted data such as school and student response rates, exclusion rates and eligibility problems.
- Translation: information relating to the verification and translation process.
- PISA Quality Monitor co-ordination: information relating to the recruitment and selection of PISA Quality Monitors and National Quality Monitoring issues.
- Data cleaners: issues identified during the data cleaning checks and from data cleaners' reports.
- Item developers: issues identified in the coder query service and training of coders.
- Data processing: issues relating to the eligibility of students tested.
- Questionnaire data: issues relating to the questionnaire data in the national questionnaire reports provided by the international contractor.
- Questionnaire Final Optical Check: issues arising from the Final Optical Check of the questionnaires.

There were two types of PISA quality monitoring 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 PISA Quality Monitoring reports contain data related to Test Administration as well as a record of interview with School Co-ordinators. 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 recorded by the National Centre and submitted as part of the national dataset to the international contractor. The National Centre quality monitor interview schedule contained information on all the standards, as did the Main Survey review.

The *National Centre Quality Monitor Interview Schedule* and the *Main Survey Review* were self-declared by the National Project Manager. The PISA Quality Monitoring data are collected independently of the National Project Manager.

Data adjudication process

The main aim of the adjudication process is to make a judgement 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 administration period.

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 international contractor staff to address any identified areas of concern. Where necessary, the relevant National Project Manager was contacted through the international 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 international contractor and the Technical Advisory Group reviewed the reports and made a determination with regard to the quality of the data from each adjudicated entity.

Monitoring compliance to any single standard occurs through responses to one or more quality assurance questions regarding test implementation and national procedures which may come from more than one area. For example, the session report data are used in conjunction with the PISA quality monitoring reports and information from the adaptation of national manuals to assess compliance with the PISA session timing standard (Standard 6.1, Annex F).

Information was collected in relation to these standards through a variety of mechanisms: through PISA Quality Monitor reports; through the Field Trial and Main Survey reviews; through information negotiated and stored on the MyPISA website (the portal which was used in PISA 2012) in relation to specific PISA implementation tasks; through communications and visits of international contractor staff to National Centres; through the formal and informal exchanges between the international contractor and National Centres over matters such as sampling, translation and verification, specially requested analyses (such as non-response bias analysis); through a detailed post-hoc inspection of all Main Survey assessment materials (test booklets); and through the data cleaning and data submission process.

For PISA 2012, an adjudication database was developed to capture, summarise and store the most important information derived from these various information sources. The staff members of the international contractor who lead each area of work were responsible for identifying relevant information, and entering it into the database. This means that at the time of data adjudication, relevant information is easily accessible for making recommendations about the fitness of use of data from each PISA adjudicated entity.

The adjudication database captures information related to the major phases of the data operation: field operations, sampling, computer-based problem solving and computer-based assessment of literacy (where applicable), questionnaires, cognitive tests. Within each of these phases, the specific activities are identified, and linked directly to the corresponding standards.

Within each section of the database, specific comments are entered that describe the situation of concern, the source of the evidence about that situation, and the recommended action. Each entry is classified as serious, minor or is rated as being of no importance for adjudication. Typically, events classified as serious would warrant very close expert scrutiny, and possibly action affecting adjudication outcomes. For example, the reliability of parental occupation data from Albania was subject to scrutiny, resulting in a recommendation that all data dependant on Albania's parental occupation data (in particular, all data that use the *PISA index of economic, social and cultural status* [ESCS]) should be deleted from the database and relevant tables. Events classified as minor would typically not directly affect adjudication outcomes, but will be reported back to National Centres to assist them in reviewing their national procedures. 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 National Project Manager, 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; and
- that the data not be endorsed for inclusion in the PISA 2012 Database.

Throughout PISA 2012, the international contractor 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.

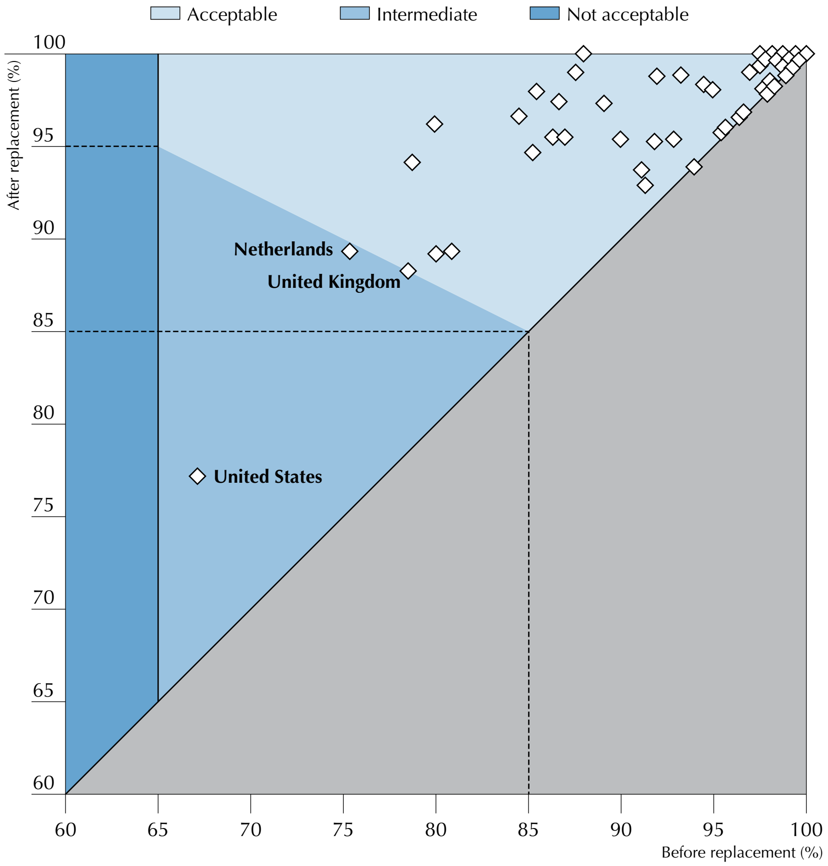
GENERAL OUTCOMES

Overview of response rate issues

The PISA school response rate requirements are discussed in Chapter 4. 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 light blue shaded region were regarded as fully satisfying the PISA school response rate criterion.

■ Figure 14.1 ■

Attained school response rates



One country – the United States – failed to meet the school response rate requirements. One other country, the Netherlands, had a response rate below the 85% level before the replacement but cleared the acceptable level after the replacement schools were included.

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

The international contractor determined that the data were acceptable.

Computer-Based Assessment

The PISA 2012 Technical Standards did not include any specification of the response rate needed for students undertaking the computer-based assessment. In the absence of agreed technical standards for the response rate of students undertaking the computer-based assessment, the Technical Advisory Group advised for the PISA 2009 Digital Reading Assessment that the desired response rate was 0.8 of the response rate of students undertaking the paper-based assessment. This response rate requirement has been carried into PISA 2012. Comments for the response rates of the countries that implemented the computer-based assessment are discussed based on those criteria.

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 Technical Advisory Group meeting that took place in Melbourne in April 2013, the quality assurance and monitoring processes outlined earlier in this chapter and in Chapter 7 had been implemented. Data adjudication focused on residual issues that remained after these quality assurance processes had been carried out.

The remaining 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 international contractor experts in advance of the assessment and alternative approaches were considered jointly between the National Project Manager and the international contractor. In isolated departures from best practice in cases such as these, a judgement might easily be made by international contractor 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 Technical 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 4 500 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 were usually available to suggest the data were suitable for use.

Particular attention has been paid to the achievement of the specified response rates of 85% for schools, 80% for students within schools and no more than 5% of students excluded from the assessment.

Albania

Analysis of the data for Albania suggest that the PISA Technical Standards may not have been fully met for the following four main reasons: *(i)* the rate of missing data for parental occupation was systematically related to (sampling) stratum membership of each student (students from higher strata had more missing data); *(ii)* 80.3% of the School Questionnaires returned contained impossible total school enrolment data, as there were more students in the Student Questionnaire data file than the number of students listed as enrolled at the school; *(iii)* the coding of items in Albania is at an extremely high level of agreement between independent coders; and *(iv)* the expected relationships between student achievement and student background characteristics were inconsistent with previous experience. Further investigation of the survey instruments, the procedures for test implementation and coding of student responses at the national level provided sufficient evidence of systematic errors and violations of the PISA Technical Standards for the parental occupation data and school enrolment data. Albania's data for parental occupation (*ESCS*, *HISEI*, *BMMJ* and *BMFJ*) and school enrolment (*SC07Q01*, *SC07Q02*, *SCHSIZE*, *PCGIRLS*, *STRATIO*, *SMRATIO* and *IRATCOMP*) are, therefore, deleted from the PISA 2012 international dataset.

Argentina

CABA (Ciudad Autónoma de Buenos Aires)

There were fewer than the 1500 students specified in the standards for an adjudicated region assessed (1316) and fewer than 50 schools participating (48). Data were included in the final database.

Belgium

Belgium had a school response rate before replacements of 83.67%. After replacement the response rate was 95.92% which was above the PISA standard.

Flemish Community

The Flemish Community had a school response rate before replacements of 79.56%. After replacement the response rate was 95.58% which was above the PISA standard.

Canada

There was a total of 6.38% exclusions in Canada. A bias analysis showed that the non-response bias would be negligible. It was thought that the extra students excluded were special needs students.

Canada's data were, therefore, included in the final database.

Denmark

Overall exclusions were greater than 5% (6.18%). Data were fully explained – there was a difficulty in defining the school population – some international schools were not included when they should have been. Denmark also had a school response rate before replacements of 84.97%. After replacement the response rate was 92.62% which was above the PISA standard. Denmark's data were included in the international database.

Estonia

There was a total of 5.80% exclusions in Estonia. A bias analysis showed that the non-response bias would be negligible. It was thought that the extra students excluded were special needs students.

Estonia's data were, therefore, included in the final database.

Hong Kong-China

Hong Kong had a school response rate before replacements of 78.85%. After replacement the response rate was 94.23% which was above the PISA standard. Fewer than 150 schools participated (147), but this was deemed to be acceptable and Hong Kong's data were included in the international database.

Iceland

Fewer than 150 schools (133) participated for Iceland. This was deemed to be acceptable, as Iceland was a full census country for the paper-based assessment, and Iceland's data were included in the international database.

Luxembourg

There was a total of 8.40% exclusions in Luxembourg. Further analysis indicated that the non-response bias would be negligible. The data from Luxembourg, therefore, were included in the international database.

Netherlands

The Netherlands had a school response rate before replacements of 74.37%. After replacement the response rate was 88.94% which was above the PISA standard. There were also fewer than 4500 students assessed (4434). Data were included in the international database.

New Zealand

New Zealand had a school response rate before replacements of 79.19%. After replacement the response rate was 89.85% which was above the PISA standard. Data were included in the international database.

Norway

There was a total of 6.11% exclusions in Norway. Data were included in the final database.

Poland

Poland had a school response rate before replacements of 84.57%. After replacement the response rate was 96.81% which was above the PISA standard. Data were included in the international database.

Qatar

The tests were incorrectly printed. After completing the Final Optical Check the printer made changes which resulted in questions being presented to the students in non-standard ways. Item difficulty was calculated and no systematic influence was observed in these cases. The data from Qatar, therefore, were included in the international database.

Spain

Balearic Islands

There was a total of 7.41% exclusions, and fewer than 1500 students assessed (1455) in the Balearic Islands. Data were included in the final database.

Catalonia

There was a total of 6.42% exclusions in Catalonia. Data were included in the final database.

Aragon

There were fewer than 1500 students assessed (1427). Data were included in the final database.

Murcia

Murcia had fewer than 1500 students assessed (1411). Data were included in the final database.

Sweden

There was a total of 5.44% exclusions in Sweden. Data were included in the final database.

Tunisia

Tunisia had fewer than 4500 students assessed (4391). Data were included in the final database.

United Kingdom

There was a total of 5.49% exclusions in the United Kingdom. Data were included in the final database.

The United Kingdom had a school response rate before replacements of 84.44%. After replacement the response rate was 90.16% which was above the PISA standard.

United States

There was a total of 5.35% exclusions in the United States. Additional analysis supported the case that no notable bias would result from non-response. It was thought that the extra students excluded were special needs students.

The United States had a school response rate of 77.78%. Additional analysis supported the case that no notable bias would result from non-response. The data from the United States, therefore, were included in the international database.

Florida

There was a total of 8.29% exclusions in Florida. Data were included in the final database.

Massachusetts

There were fewer than 50 schools assessed (49). Data were included in the final database.

Note

1. Not all regions opt to undergo the full adjudication that would allow their results to be compared statistically to all other participating economies and adjudicated regions. For example, the states of Australia are not adjudicated regions whereas the Flemish Community of Belgium is an adjudicated region.

15

Proficiency Scale Construction

INTRODUCTION

This chapter discusses the methodology used to develop the PISA reporting scales that describe a number of levels of proficiency in the different PISA literacy variables, and presents the outcomes of that development process.

For many years, the Australian Council for Educational Research (ACER) has used and progressively refined an approach to substantive interpretation of scales based on item calibration, employing a reporting mechanism generally known as "described proficiency scales", alternatively referred to more recently as "learning metrics", as part of its analysis and reporting of test results. The approach has its origins in work of Benjamin Wright and his collaborators at the University of Chicago from the 1960s. An early published example of a dimension laid out using Rasch-based item calibrations and illustrated with the items and their characteristics is found in Wright and Stone (1979). A similar approach has been used in a number of Australian assessment projects, dating back at least to the TORCH project that originated in Western Australia in 1982-1983 and was published later by ACER (Mossenson, Hill and Masters, 1987), the Basic Skills Testing Programme in New South Wales in 1989 (Masters et al., 1990), as well as in many more recent projects. ACER has used the approach in the reporting of PISA results from its inception: for two administrations in which reading literacy was the major test domain, two in which mathematics was the major domain, and one in which science took centre stage. The same approach was also used to report problem solving in 2003 and 2012, digital reading in 2009, and financial literacy in 2012. Reporting the results of the National Assessment of Educational Progress (NAEP) in the United States since 1990 has also used substantive descriptions of typical accomplishments at points along their reporting scales (see Bourque, 2009) for a history of NAEP reporting), using a consensus-based approach along with 'scale anchoring' to define levels and cut-points (Beaton and Allen, 1992).

This chapter presents the methodology, and the products of the application of the methodology, for reporting of PISA 2012 survey outcomes. PISA reports student performance not just as numerical scores, but also in terms of content, by describing what students who achieve a given level on a PISA scale typically know and can do. This chapter explains how these described proficiency scales are developed, and also how the results are reported and how they can be interpreted.

PISA has adopted an approach to reporting survey outcomes that involves the development of learning metrics, which are dimensions of educational progression. A learning metric is usually depicted as a line with numerical gradations that quantify how much of the measured variable is present. Locations along this metric can be specified by numerical 'scores', or can be described substantively, hence the label for these metrics used in PISA: described proficiency scales. The scales are called "proficiency scales" rather than "performance scales" because they report what students typically know and can do at given levels, rather than what the individuals who were tested actually did on a single occasion (the test administration). This is because PISA is interested in reporting general results, rather than the results of individuals. PISA uses samples of students and items to make estimates about populations: a sample of 15-year-old students is selected to represent all the 15-year-olds in a country, and a sample of test items from a large pool is administered to each student. Results are then analysed using statistical models that estimate the likely proficiency of the population, based on this sampling.

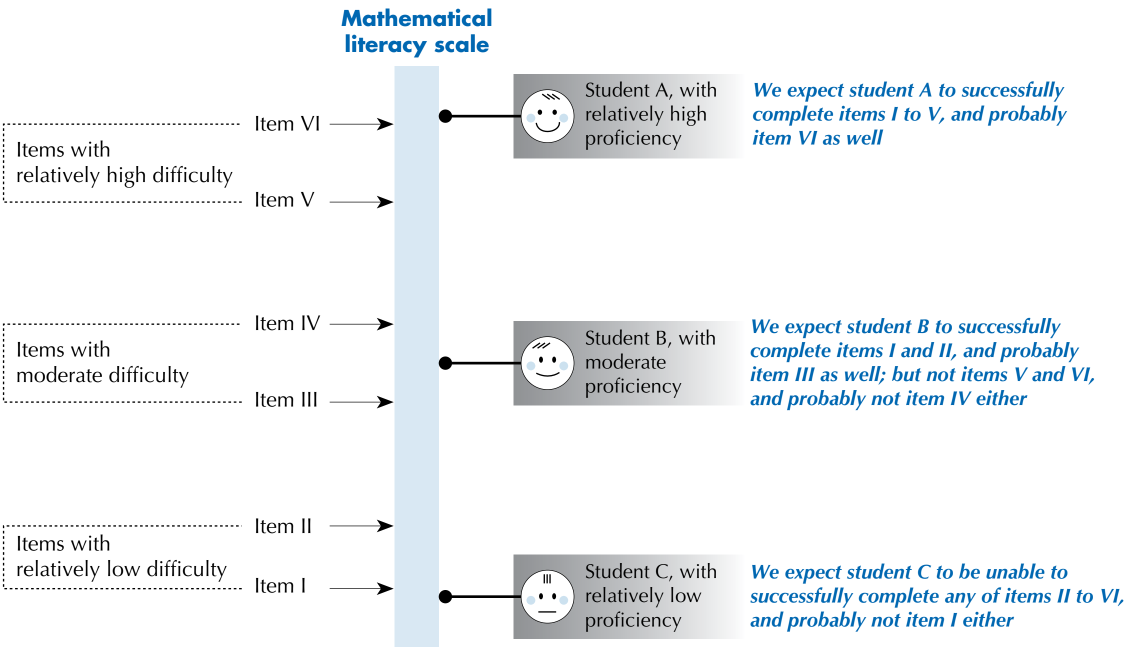
The PISA test design makes it possible to use techniques of modern item response modelling (see Chapter 9) to simultaneously estimate the ability of all students taking the PISA assessment, and the difficulty of all PISA items, locating these estimates of student ability and item difficulty on a single continuum. In this context, the single continuum is a way to represent the variable of interest – the "student ability" is determined by the extent to which a student possesses the key components of the variable, and the "item difficulty" is determined by the extent to which responding to the item requires activation of the variable.

The relative ability of students taking a particular test can be estimated by considering the proportion of test items to which they provide a correct response, and the difficulty of the items. The relative difficulty of items in a test can be estimated by considering the proportion of test takers getting each item correct, and the ability of the students. The mathematical model employed to analyse PISA data, generated from a rotated test design in which students take different but overlapping tasks, is implemented through test analysis software that uses iterative procedures to simultaneously estimate the likelihood that a particular person will respond correctly to a given test item, and the likelihood that a particular test item will be answered correctly by a given student. The result of these procedures is a set of estimates that enables a continuum (the learning metric) to be defined, which is a realisation of the variable of interest. On that continuum it is possible to estimate the location of individual students, thereby seeing how much of the variable of interest they demonstrate, and it is possible to estimate the location of individual test items, thereby seeing how much of the variable each item embodies. This continuum is referred to as the overall PISA literacy scale in the relevant test domain (such as reading, mathematics or science).

PISA assesses students, and uses the outcomes of that assessment to produce estimates of students' proficiency in relation to a number of literacy variables. These variables are defined in the relevant PISA literacy framework (OECD, 2013). For each of these literacy variables, one or more scales are defined, which stretch from very low levels of literacy through to very high levels. What such a scale means in terms of student proficiency is that students whose ability estimate places them at a certain point on the PISA literacy scale would most likely be able to successfully complete tasks at or below that location, and increasingly more likely to complete tasks located at progressively lower points on the scale, but would be less likely to be able to complete tasks above that point, and increasingly less likely to complete tasks located at progressively higher points on the scale. Figure 15.1 depicts a literacy scale, stretching from relatively low levels of literacy at the bottom of the figure, to relatively high levels towards the top. Six items of varying difficulty are placed along the scale, as are three students of varying ability. The relationship between the students and items at various levels is described.

■ Figure 15.1 ■

The relationship between items and students on a proficiency scale



It is possible to describe the scales using words that encapsulate various demonstrated competencies typical of students possessing varying amounts of the underlying literacy constructs. Each student's location on those scales is estimated, and those location estimates are then aggregated in various ways to generate and report useful information about the literacy levels of 15-year-old students within and among participating countries.

Development of the details of the method of describing proficiency in PISA reading, mathematical and scientific literacy occurred in the lead-up to the reporting of outcomes of the PISA 2000 survey and was revised in the lead-up to the PISA 2003, 2006 and 2009 surveys. Essentially, the same methodology has again been used to develop proficiency descriptions for PISA 2012. Given the volume and breadth of data that were available from the PISA 2012 assessment when mathematics was the major assessment domain, review and extension of the descriptions of mathematical literacy that had been developed from the PISA 2003 data became possible. The detailed proficiency descriptions that had been developed for the reading domain in PISA 2009 were used again, and the descriptions used for science in 2006 were used again, in both cases with the reduced data available from the 2012 administration in which those were minor assessment domains. In addition, new described proficiency scales for problem solving and for financial literacy were developed.

The Mathematics Expert Group worked with the PISA international contractor to review and revise the sets of described proficiency scale and subscales for PISA mathematics. Similarly, the international contractor worked with the Problem Solving and Financial Literacy Expert Groups to develop the described proficiency scales for these domains.

DEVELOPMENT OF THE DESCRIBED SCALES

Since PISA 2000, the development of described proficiency scales for PISA has been carried out through a process involving a number of stages. The stages are described here in a linear fashion, but in reality the development process involved some backwards and forwards movement where stages were revisited and descriptions were progressively refined.

Stage 1: Identifying possible scales

The first stage in the process involved the experts in each domain articulating possible reporting scales (dimensions) for the domain.

In the case of mathematics, a single proficiency scale was originally developed for PISA 2000. With the additional data available in the 2003 survey cycle, when mathematics was the major test domain for the first time, the possibility of reporting according to the four overarching ideas or the three competency clusters described in the PISA mathematics framework applicable at that time were both considered. Accordingly, in 2003 subscales based on the four overarching ideas – *space and shape*, *change and relationships*, *quantity* and *uncertainty* – were reported. In PISA 2006 and PISA 2009, when mathematics was again a minor domain, a single mathematics scale only was reported.

For PISA 2012, a review of the reporting structure for mathematics was carried out by the expert group as part of a comprehensive revision of the framework, in conjunction with ACER staff, and at the specific behest of the PISA Governing Board that had indicated clearly that it was interested in seeing mathematical process dimensions used as the primary basis for reporting in mathematics. As well as considering ways in which this could be done, the mathematics expert group also had to consider how the addition of an optional computer-based assessment component could be incorporated in the reporting of the PISA mathematical outcomes. The result of these considerations was firstly, that the computer-based items would be used to expand the scope of expression of the same mathematical literacy dimension that is expressed through the paper-based items; and secondly that the reporting of three process-based subscales labelled *formulating situations mathematically* (usually abbreviated to "formulate"), *employing mathematical concepts, facts, procedures and reasoning* (usually abbreviated to "employ"), and *interpreting, applying and evaluating mathematical outcomes* (with the abbreviation "interpret") would be supported. In addition, for continuity with the PISA 2003 reporting scales, the content-based scales were also reported, with the labels *space and shape*, *change and relationships*, *quantity* and *uncertainty and data* (the latter being the same dimension as the previous *uncertainty* subscale, but with a new label).

For reading in the PISA 2000 survey cycle, two main options were actively considered – scales based on the type of reading task, and scales based on the form of reading material. For the international report, the first of these was implemented, leading to the development of scales to describe the types of reading tasks, or "aspects" of reading: a subscale for *retrieving information*, a second subscale for *interpreting texts* and a third for *reflection and evaluation*. The thematic report for PISA 2000, *Reading for Change*, also reported on the development of subscales based on the form of reading material: *continuous texts* and *non-continuous texts* (OECD, 2002). Volume I of the *PISA 2009 Results* included descriptions of both of these sets of subscales as well as a combined print reading scale (OECD, 2010). The names of the aspect subscales were modified in order to better apply to digital as well as print reading tasks. The modified aspect category names are *access and retrieve* (replacing *retrieving information*), *integrate and interpret* (replacing *interpreting texts*) and *reflect and evaluate* (for *reflection and evaluation*). For digital reading, a separate, single scale was developed based on the digital reading assessment items administered in 19 countries in PISA 2009 as an international option (OECD, 2011). For PISA 2012, when reading reverted to minor domain status, a single print reading scale was reported, along with a single digital reading scale.

For science, given the small number of items in PISA 2000 and 2003, a single overall proficiency scale was developed to report results. As with mathematics in 2003, the expanded focus on science in 2006 allowed for a division into scales for reporting purposes. Two forms of scale were considered. One of these was based on definitions of scientific competencies involving the identification of scientific issues, the explanation of phenomena scientifically and the use of

scientific evidence. The other form separated scientific knowledge into “knowledge of science” involving the application of scientific concepts in the major fields of physics, chemistry, biology, earth and space science, and technology; and “knowledge about science” involving the central processes underpinning the way scientists go about obtaining and using data – in other words, understanding scientific methodology. The scales finally selected for inclusion in the PISA 2006 database were the three competency-based subscales: *identifying scientific issues*, *explaining phenomena scientifically* and *using scientific evidence* (OECD, 2007). In PISA 2009 and PISA 2012, science as a minor domain was reported as a single scale only.

Wherever subscales were under consideration, they arose clearly from the framework for the domain, they were seen to be meaningful and potentially useful for feedback and reporting purposes, and they needed to be defensible with respect to their measurement properties. Due to the longitudinal nature of the PISA project, the decision about the number and nature of reporting scales also had to take into account the fact that in some test cycles a domain will be treated as minor and in other cycles as major.

For problem solving, and for the optional assessment component of financial literacy, in both of which a rather limited volume of data were available based on a relatively small number of test items, proficiency descriptions of a single overall dimension were developed in each domain.

Stage 2: Assigning items to scales

The second stage in the process was to associate each test item used in the study with each of the subscales under consideration. Domain experts (including members of the relevant subject matter expert group, the test developers and staff of the international contractor) judged the characteristics of each test item against the relevant framework categories.

Stage 3: Skills audit

The next stage involved a detailed expert analysis of each item, and in the case of items with partial credit, for each score step within the item, in relation to the definition of the relevant subscale from the domain framework. The skills and knowledge required to achieve each score step were identified and described.

This stage involved negotiation and discussion among the experts involved, circulation of draft material, and progressive refinement of drafts on the basis of expert input and feedback. Further detail on this analysis is provided below.

Stage 4: Analysing Field Trial data

For each set of scales being considered, the Field Trial item data were analysed using item response techniques to derive difficulty estimates for each achievement threshold for each item.

Many items had a single achievement threshold (associated with students providing a correct rather than incorrect response). Where partial credit was available, more than one achievement threshold could be calculated (achieving a score of one or more rather than zero, two or more rather than one, and so on).

Within each scale, achievement thresholds were placed along a difficulty continuum linked directly to student abilities. This analysis gives an indication of the utility of each scale from a measurement perspective.

Stage 5: Defining the dimensions

The information from the domain-specific expert analysis (Stage 3) and the statistical analysis (Stage 4) were combined. For each set of scales being considered, the item score steps were ordered according to the magnitude of their associated thresholds and then linked with the descriptions of associated knowledge and skills, giving a hierarchy of knowledge and skills that defined the dimension. Clusters of skills were found using this approach, which provided a basis for understanding each dimension and describing proficiency in different regions of the scale.

Stage 6: Revising and refining with Main Survey data

When the Main Survey data became available, the information arising from the statistical analysis about the relative difficulty of item thresholds was updated. This enabled a review and revision of Stage 5. The preliminary descriptions and levels were then reviewed and revised, and the approach to defining levels and associating students with those levels that had been used in the reporting of PISA 2000, 2003, 2006 and 2009 results was applied.

DEFINING AND INTERPRETING PROFICIENCY LEVELS

How should we divide the proficiency continuum up into levels that might have some utility? And having defined levels, how should we decide on the level to which a particular student should be assigned? What does it mean to be at a level?

The relationship between the student and the items is probabilistic: that is, there is some probability that a particular student can correctly answer any particular item. If a student is located at a point above an item, the probability that the student can successfully complete that item is relatively high, and if the student is located below the item, the probability of success for that student on that item is relatively low. This leads to the question as to the precise criterion that should be used to locate a student on the same scale as that on which the items are laid out. When placing a student at a particular point on the scale, what probability of success should we deem sufficient in relation to items located at the same point on the scale? If a student were given a test comprising a large number of items each with the same specified difficulty, what proportion of those items would we expect the student to successfully complete? Or, thinking of it in another way, if a large number of students of equal ability were given a single test item having a specified item difficulty, about how many of those students would we expect to successfully complete the item?

The answer to these questions is essentially arbitrary, but in order to define and report PISA outcomes in a consistent manner, we need an approach to defining performance levels, and to associating students with those levels. This is both a technical and very practical matter of interpreting what it means to be at a level, and has very significant consequences for reporting national and international results. The methodology that was developed and used for PISA 2000, 2003, 2006 and 2009 was essentially retained for PISA 2012.

Several principles were considered for developing and establishing a useful meaning for being at a level, and therefore for determining an approach to locating cut-off points between levels and associating students with them. The overriding need to develop and promote a common understanding of the meaning of levels was recognised. First, it is important to understand that the literacy skills measured in PISA must be considered as continua: there are no natural breaking points to mark borderlines between stages along these continua. Dividing each of these continua into levels, though useful for communication about students' development, is essentially arbitrary. Like the definition of units on, for example, a scale of length, there is no fundamental difference between 1 metre and 1.5 metres – it is a matter of degree. It is useful, however, to define stages, or levels along the continua, because they enable us to communicate about the proficiency of students in terms other than numbers. The approach adopted for PISA 2000 was that it would only be useful to regard students as having attained a particular level if this would mean that we can have certain expectations about what these students are capable of in general when they are said to be at that level. It was decided that this expectation would have to mean at a minimum that students at a particular level would be more likely than not to successfully complete tasks at that level. By implication, it must be expected that they would succeed on at least half of the items on a test composed of items uniformly spread across that level. This definition of being "at a level" is useful in helping to interpret the proficiency of students at different points across the proficiency range defined at each level.

For example, students at the bottom of a level would complete at least 50% of tasks correctly on a test set at the level, while students at the middle and top of each level would be expected to achieve a higher success rate. At the top end of the bandwidth of a level would be the students who have mastered that level. These students would be likely to solve a high proportion of the tasks at that level. But, being at the top border of that level, they would also be at the bottom border of the next level up, where according to the reasoning here they should have a likelihood of at least 50% of solving any tasks defined to be at that higher level.

Further, the meaning of being at a level for a given scale should be more or less consistent for each level, indeed also for scales from the different domains. In other words, to the extent possible within the substantively based definition and description of levels, cut-off points should create levels of more or less constant breadth. Some small variation may be appropriate, but in order for interpretation and definition of cut-off points and levels to be consistent, the levels have to be about equally broad within each scale. Clearly this would not apply to the highest and lowest proficiency levels, which are unbounded.

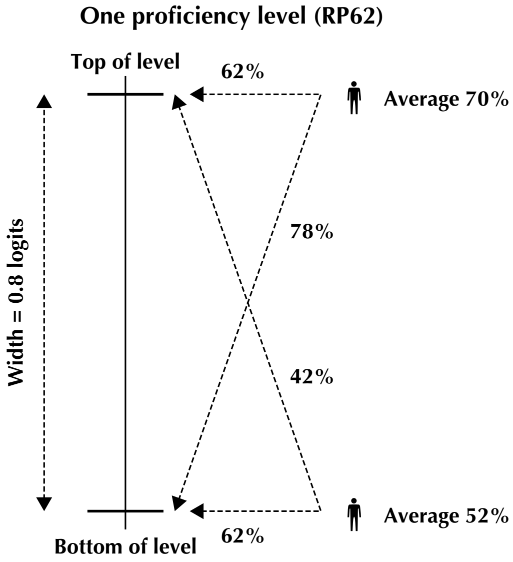
A more or less consistent approach should be taken to defining levels for the different scales. Their breadth may not be exactly the same for the proficiency scales in different domains, but the same kind of interpretation should be possible for each scale that is developed. The approach links the two variables mentioned in the preceding paragraphs, and a third related variable. The three variables can be expressed as follows:

- the expected success of a student at a particular level on a test containing items at that level (proposed to be set at a minimum that is near 50% for the student at the bottom of the level, and higher for other students in the level);
- the width of the levels in that scale (determined largely by substantive considerations of the cognitive demands of items at the level and observations of student performance on the items); and
- the probability that a student in the middle of a level would correctly answer an item of average difficulty for that level (in fact, the probability that a student at any particular level would get an item at the same level correct), sometimes referred to as the "RP-value" for the scale (where "RP" indicates "response probability").

Figure 15.2 summarises the relationship among these three mathematically linked variables under a particular scenario. The vertical line represents a segment of the proficiency scale, with marks delineating the "top of level" and "bottom of level" for any level one might want to consider, with a width of 0.8 logits between the boundaries of the level (but note that this width can vary somewhat for different domains). The RP62 indicates that any person will be located on the scale at a point that gives him or her a 62% chance of getting an item at that same level correct. The person represented near the top of the level shown has a 62% chance of getting an item correct that is located at the top of the level, and similarly the person represented at the bottom of the level has the same chance of correctly answering a question at the bottom of the level. A person at the bottom of the level will have an average score of about 52% correct on a set of items spread uniformly across the level. Of course that person will have a higher likelihood (62%) of getting an item at the bottom of the level correct, and a lower likelihood (about 42%) of getting an item at the top of the level correct. A person at the top of the level will have an average score of about 70% correct on a set of items spread uniformly across the level. Of course that person will have a higher likelihood (about 78%) of getting an item at the bottom of the level correct, and a lower likelihood (62%) of getting an item at the top of the level correct.

■ Figure 15.2 ■

Calculating the RP value used to define PISA proficiency level



PISA 2000 implemented the following solution that was then used in all subsequent survey administrations: start with the range of described abilities for each bounded level in each scale (the desired band breadth); then determine the highest possible RP value that will be common across domains potentially having bands of slightly differing breadth that would give effect to the broad interpretation of the meaning of being at a level (an expectation of correctly responding to a minimum of 50% of the items in a test comprising items spread uniformly across that level). The value RP=0.62 is a probability value that satisfies the logistic equations through which the scaling model is defined, subject to the two constraints mentioned earlier (a width per level of about 0.8 logits and the expectation that a student would get at

least half of the items correct on a hypothetical test composed of items spread evenly across the level). In fact RP=0.62 satisfies the requirements for any scales having band widths up to about 0.97 logits.

With the constraint of a minimum 50% mentioned above, which is central to the definition of the PISA proficiency levels, the RP value required for scales composed of bands of other widths is given by the equation in Figure 15.2, where *x* is the width of the bands.

$$RP \geq \frac{\exp\left(x/2\right)}{1+\exp\left(x/2\right)}$$

The highest and lowest levels are unbounded. For a certain high point on the scale and below a certain low point, the proficiency descriptions could, arguably, cease to be applicable. At the high end of the scale, this is not such a problem since extremely proficient students could reasonably be assumed to be capable of at least the achievements described for the highest level. At the other end of the scale, however, the same argument does not hold. A lower limit therefore needs to be determined for the lowest described level, below which no meaningful description of proficiency is possible. It was proposed that the floor of the lowest described level be set so that it was the same breadth as the other bounded levels. Student performance below this level is lower than that which PISA can reliably assess and, more importantly, describe.

REPORTING THE RESULTS FOR PISA MATHEMATICS

In this section, the way in which levels of mathematical literacy are defined, described and reported will be discussed. They will be exemplified using a number of items from the PISA 2012 assessment. The mathematics scale and content subscales were developed from the corresponding scale and subscales established in PISA 2003 (OECD, 2004), whereas the process subscales were created as a completely new measure.

Building an item map for mathematics

The data from the PISA mathematics assessment were processed to generate a set of item difficulty measures initially for the 290 paper-based and computer-based items used in the Field Trial that took place in 2011, and reviewed using the 150 items included in the Main Survey. In fact, when the difficulty measures that were estimated for each of the partial credit steps of the polytomous items are also taken into account, 168 item difficulty estimates were generated from the Main Survey items.

The 6-step analysis of items described earlier was carried out as the mathematics items were developed. This analysis included judgements about the elements of the PISA mathematics framework that were relevant to each item. For example, each item was analysed to determine which of the newly defined process categories was most significantly involved in a successful response.

Following data analysis and the resultant generation of difficulty estimates for each of the 168 item steps (and the additional item steps from the Field Trial items), the items and item steps were associated with their difficulty estimates, with their framework classifications, and with their brief qualitative descriptions. Figure 15.3 shows a map of some of this information from a sample of items from the PISA 2012 test, the items that made up two complete clusters in the test that were released publicly following the release of PISA 2012 results. Each row in Figure 15.3 represents an individual item or item step. The selected items and item steps have been ordered according to their difficulty, with the most difficult of these steps at the top, and the least difficult at the bottom. The difficulty estimate for each item and step is given in PISA scale units, along with the associated classifications and descriptions.

When a map such as this is prepared using all available items, it becomes possible to look for factors that are associated with item difficulty. This can be done by referring to the ways in which mathematical literacy is associated with questions located at different points ranging from the bottom to the top of the scale. For example, the item map in Figure 15.3 shows that the easiest items tend to involve identifying mathematical information presented in a table or graph and linking that information to some element of the problem context. The most difficult items, by contrast, are based on knowledge of particular mathematical content or procedures, and they involve several steps that require some creativity or strategic control in linking the context to the mathematical representation of aspects of the context, and often substantial mathematical processing or calculation to devise a solution.

■ Figure 15.3 ■

A map for selected mathematics items

Item Code	Item Name	Item difficultiy on PISA scale	Description of item demand	Process: Formulate	Process: Employ	Process: Interpret	Content: Change and relationships	Content: Quantity	Content: Space and shape	Content: Uncertainty and data
PM995Q02	Revolving Door Q2	840.3	Apply knowledge of circle geometry and reasoning to interpret a given geometric model and to formulate it mathematically enabling a solution	●					●	
PM923Q04	Sailing Ships Q4	702.1	Devise and implement a multi-step strategy involving significant modelling and extended calculation to formulate then solve a complex real world problem involving fuel costs and volume, equipment costs	●			●			
PM957Q03	Helen the Cyclist (E) Q3	696.6	Interpret information about distance and speed, devise a representation to help formulate a model for average speed, calculate average speed including converting units		●		●			
PM991Q02.2	Garage Q2.2	687.3	Interpret task demand from text and diagrams, formulate area calculation process from given measurements and specification (correct working and justification)		●				●	
PM991Q02.1	Garage Q2.1	663.2	Interpret task demand from text and diagrams, formulate area calculation process from given measurements and specification (partially correct result)		●				●	
PM903Q01.2	Drip Rate Q1.2	657.7	Interpret text and equation linking four variables, provide explanation of effect of specified change to one variable on a second variable if all other variables remain unchanged		●		●			
PM942Q02	Climbing Mount Fuji Q2	641.6	Follow multi-step strategy to interpret information, formulate and use a model that connects given time, speeds, and distance, and implement a time calculation	●			●			
PM903Q03	Drip Rate Q3	631.7	Interpret formula linking three variables in medical context, check consistency of units, substitute two values into given equation, transpose equation and solve		●		●			
PM903Q01.1	Drip Rate Q1.1	610.5	Interpret text and equation linking four variables, provide partial explanation of effect of specified change to one variable on a second variable if all other variables remain unchanged		●		●			
PM942Q03.2	Climbing Mount Fuji Q3.2	610.0	Identify and mathematise the defined task goal; use the model to calculate an average from given data in context, in specified units		●			●		
PM934Q01	London Eye Q1	592.3	Interpret text and diagram to form a strategy: identify, extract and use data from geometric sketch to formulate a model, apply it to calculate a length		●				●	
PM942Q03.1	Climbing Mount Fuji Q3.1	591.3	Identify and mathematise the defined task goal; use the model to calculate an average from given data in context, in specified units (answer correct but expressed in wrong units)		●			●		
PM00FQ01	Apartment Purchase Q1	576.2	Interpret graphic representation, use geometric reasoning to identify relevant dimensions needed to carry out specified area calculation with several components	●					●	
PM995Q03	Revolving Door Q3	561.3	Use reasoning to formulate and apply a proportional model involving several steps	●				●		
PM985Q03	Which Car? Q3	552.6	Interpret information on tax rate for a purchase to formulate a simple model, locate and extract data from table, and calculate a percentage		●			●		
PM923Q03	Sailing Ships Q3	538.5	Use geometry knowledge (trigonometry, or Pythagoras) to form a simple model to solve a right-angled triangle in context, evaluate and select answer from given options		●				●	
PM923Q01	Sailing Ships Q1	511.7	Interpret text and quantitative information; use reasoning and calculation to implement a percentage increase, and select from given options		●			●		
PM957Q02	Helen the Cyclist Q2	510.6	Interpret information about distance and speed, devise a simple proportional model to calculate a time corresponding to given distance and average speed		●		●			
PM985Q02	Which Car? Q2	490.9	Identify smallest of four decimal numbers from data table, use place value in context		●			●		
PM924Q02	Sauce Q2	489.1	Follow a multi-step strategy to devise and apply a suitable proportional model and perform the resultant percent calculation	●				●		
PM934Q02	London Eye Q2	481.0	Interpret text to understand task, extract and use data from graphic to formulate simple model, involving reasoning about fractions of a circle	●					●	
PM942Q01	Climbing Mount Fuji Q1	464.0	Interpret text to understand task; formulate strategy - define a time period in required unit (days), and combine information to devise a method to calculate a daily average; perform the calculation	●				●		
PM957Q01	Helen the Cyclist Q1	440.5	Interpret information about the distance travelled in two time periods to verify a given conclusion about the corresponding average speeds		●		●			
PM918Q05	Charts Q5	428.2	Identify and extract relevant data from a bar graph, model trend and use it to interpolate		●					●
PM991Q01	Garage Q1	419.6	Use spatial reasoning: devise a comparison strategy to identify correct representational model from given options			●			●	
PM918Q02	Charts Q2	415.0	Interpret bar graph; identify and extract data value defined by comparative condition to answer a question about the context			●				●
PM918Q01	Charts Q1	347.7	Interpret bar graph, identify and extract data value to answer a question about the context			●				●
PM985Q01	Which Car? Q1	327.8	Identify data in a table meeting specifications of simple mathematical relationships			●				●

More generally, the difficulty of mathematics questions in PISA 2012 is associated with a number of item characteristics that can be seen as calling forth varying levels of activation by students of each member of the set of fundamental mathematical capabilities described in the mathematics framework. That set of capabilities has been useful in exposing the ways in which cognitive demand varies among different items, and has provided a rich means of describing different levels of proficiency.

- Mathematical *communication* involves understanding the stated task objectives and the mathematical language used, recognising what information is relevant and what is the nature of the response needed; and also may involve the active steps including some or all of presenting the response, solution steps, description of the reasoning used and justification of the answer provided. Demand for this capability increases according to the complexity of material to be interpreted in understanding the task, the need to link multiple information sources or to move repeatedly among information elements; and with the need to provide a detailed written solution or explanation.
- Item complexity and difficulty is also affected by the nature and extent of *strategic thinking* that is required to progress towards a problem solution. In the simplest problems, the solution path is specified or it is obvious, and involves perhaps just a single processing step, while in other problems a solution strategy may involve drawing on several elements of mathematical knowledge, linking them in a particular sequence of related steps, and exercising quite a degree of control to keep sight of the objective and the way the stages of a solution will lead to meeting essential sub-goals that will fit together in achieving the overall problem objective.
- PISA problems very frequently are set in some context of the kind individuals may encounter in their school, work or daily life. Contextualised problems may require the student to impose a transformation of information into a suitable mathematical form. This process of *mathematisation* lies at the heart of the mathematical process referred to as *formulating*. In the most difficult problems it can involve making simplifying assumptions, identifying relevant variables and devising a suitable way to express them mathematically, and understanding the relationships between the contextual elements and their mathematical expression. It can also involve forging links between mathematical results or mathematical information and the situation that information is intended to describe. Translating or *interpreting* mathematical results in relation to specific elements of the problem context, and validating the adequacy of the solution with respect to the context are also part of this mathematical capability.
- A widely recognised element of much mathematical work is the myriad ways in which mathematical information, relationships and processes can be expressed. Mathematical *representations* can take the form of equations, graphs, charts, tables, formulae and so on. These vary in familiarity to students, and in their complexity, and this variation can directly affect the difficulty of tasks that involve the use or construction of mathematical representations. Students may be presented with mathematical representations they must use or process in some way. Or they may be required to create or devise a representation of data, information or relationships in order to solve a problem. Representations can be simple, or more complex. Multiple representations may be involved or required in order to solve a problem, and tasks that involve linking two or more different representations tend to be more difficult.
- One of the most important drivers of item difficulty lies in the particular mathematical content knowledge that must be activated to solve problems, such as the number and nature of definitions, facts, rules, algorithms and procedures, especially the need to understand and manipulate symbolic expressions, formulae, functional relations or other algebraic expressions, but also the need to perform arithmetic calculations and to understand the formal rules that govern them. A problem that requires counting or adding small integers clearly imposes a different level of cognitive demand compared to an item that requires manipulating and solving an equation, or applying the Pythagoras theorem.
- Finally, the nature of the *reasoning* involved in solving a mathematical problem, and the degree to which mathematical *argumentation* must be understood or applied as part of the solution process contribute in important ways to item difficulty. The nature, number or complexity of elements that need to be brought together in making inferences, and the length and complexity of the chain of inferences needed are significant contributors to increased demand for activation of the *reasoning and argument* competency.

Levels of mathematical literacy

The approach to reporting used by the OECD has been defined in previous cycles of PISA and is based on the definition and description of a number of levels of literacy proficiency. Descriptions were developed to characterise typical student performance at each level. The levels were used to summarise the performance of students, to compare performances across subgroups of students, and to compare average performances among groups of students, in particular among the students from different participating countries. A similar approach has been used here to analyse and report PISA 2012 outcomes for mathematics.

For mathematics in PISA 2003, when the fully articulated PISA mathematics scale was first developed, student scores were transformed to the PISA scale, with a mean of 500 and a standard deviation of 100, and six levels of proficiency were defined and described. For PISA 2012, the new items together with link items from previous PISA survey administrations that were administered again in PISA 2012 were calibrated independently as a set and then equated with the PISA 2003 scale.

The mathematics level definitions on the PISA scale are given in Figure 15.4. The same definitions apply to the overall mathematical proficiency scales, and to each of the process-based and content-based subscales.

■ Figure 15.4 ■

Mathematical literacy performance band definitions on the PISA scale

Level	Score points on the PISA scale
6	Above 669.3
5	From 607.0 to less than 669.3
4	From 544.7 to less than 607.0
3	From 482.4 to less than 544.7
2	From 420.1 to less than 482.4
1	From 357.8 to less than 420.1
Below level 1	Below 357.8

The information about the items in each band is used to develop summary descriptions of the kinds of mathematical knowledge and understanding associated with different levels of proficiency. These summary descriptions can then be used to encapsulate typical mathematical proficiency of students associated with each level. As a set, they describe development in mathematical literacy.

The PISA 2003 proficiency descriptions have been revised and enriched using information from the new items developed for PISA 2012 including those delivered via computer, and the revised descriptions are presented in Figure 15.5. They are further described and illustrated in the first volume of the *PISA 2012 Results* (OECD, 2014a).

■ Figure 15.5 ■

Summary descriptions of the six proficiency levels on the mathematical literacy scale

Level	What students can typically do
6	At Level 6, students can conceptualise, generalise and utilise information based on their investigations and modelling of complex problem situations, and can use their knowledge in relatively non-standard contexts. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understanding, along with a mastery of symbolic and formal mathematical operations and relationships, to develop new approaches and strategies for attacking novel situations. Students at this level can reflect on their actions, and can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situation.
5	At Level 5 students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They begin to reflect on their work and can formulate and communicate their interpretations and reasoning.
4	At Level 4 students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilise their limited range of skills and can reason with some insight, in straightforward contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.
3	At Level 3 students can execute clearly described procedures, including those that require sequential decisions. Their interpretations are sufficiently sound to be a base for building a simple model or for selecting and applying simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships. Their solutions reflect that they have engaged in basic interpretation and reasoning.
2	At Level 2 students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions to solve problems involving whole numbers. They are capable of making literal interpretations of the results
1	At Level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli.

Figures 15.6, 15.7 and 15.8 provide the summary descriptions of skills and knowledge and understanding required to complete tasks located within the defined bands for the process subscales: *Formulating situations mathematically*; *Employing mathematical concepts, facts, procedures and reasoning;* and *Interpreting, applying and evaluating mathematical outcomes* respectively.

■ Figure 15.6 ■

Summary descriptions of the six proficiency levels on the mathematical process subscale *Formulating situations mathematically*

Level	What students can typically do
6	Students at or above Level 6 can typically apply a wide variety of mathematical content knowledge to transform and represent contextual information or data, geometric patterns or objects into a mathematical form amenable to investigation. At this level, students can devise and follow a multi-step strategy involving significant modelling steps and extended calculation to formulate and solve complex real world problems in a range of settings, for example involving material and cost calculations in a variety of contexts, or to find the area of an irregular region on a map; can identify what information is relevant (and what is not) from contextual information about travel times, distances and speed to formulate appropriate relationships among them; can apply reasoning across several linked variables to devise an appropriate way to present data in order to facilitate pertinent comparisons; can devise algebraic formulations that represent a given contextual situation.
5	At this level, students show an ability to use their understanding in a range of mathematical areas to transform information or data from a problem context into mathematical form. They can typically transform information from different representations involving several variables, into a form suitable for mathematical treatment. They can typically formulate and modify algebraic expressions of relationships among variables; can use proportional reasoning effectively to devise computations; they are typically able to draw together information from different sources to formulate and solve problems involving geometric objects, features and properties, or analyse geometric patterns or relationships and express them in standard mathematical terms; they can transform a given model according to changed contextual circumstances; can formulate a sequential calculation process based on text descriptions; can activate statistical concepts such as randomness, or sample, and apply probability, to formulate a model.
4	At Level 4, students show an ability to link information and data from related representations (for example, a table and a map, or a spreadsheet and a graphing tool) and apply a sequence of reasoning steps in order to formulate a mathematical expression needed to carry out a calculation or otherwise to solve a contextual problem. At this level, students can typically formulate a linear equation from a text description of a process, for example in a sales context, and can formulate and apply cost comparisons to compare prices of sale items; can identify which of given graphical representations corresponds to a given description of a physical process; they can specify a sequential calculation process in mathematical terms; they can identify geometrical features of a situation and use their geometric knowledge and reasoning to analyse a problem, for example to estimate areas or to link a contextual geometric situation involving similarity to the corresponding proportional reasoning; they can typically combine multiple decision rules needed to understand or implement a calculation where different constraints apply; and they can formulate algebraic expressions when the contextual information is reasonably straight-forward, for example to connect distance and speed information in time calculations.
3	At this level, students show an ability to identify and extract information and data from text, tables, graphs, maps or other representations, and make use of them to express a relationship mathematically, including interpreting or adapting simple algebraic expressions related to an applied context. Students at this level can transform a textual description of a simple functional relationship into a mathematical form, for example with unit costs or payment rates; can form a strategy involving two or more steps to link problem elements or to explore mathematical characteristics of the elements; can apply reasoning with geometric concepts and skills to analyse patterns or to identify properties of shapes or a specified map location, or to identify information needed to carry out some pertinent calculations, including calculations involving the use of simple proportional models and reasoning, where the relevant data and information is immediately accessible; and can typically understand and link probabilistic statements to formulate probability calculations in contexts such as in a manufacturing process, or a medical test.
2	At this level, students can understand written instructions and information about simple processes and tasks in order to express them in a mathematical form. They can typically use data presented in text or in a table (for example giving information about cost of some product or service) to formulate a computation required, such as, to identify the length of a time period, or to present a cost comparison, or to calculate an average; can analyse a simple pattern, for example by formulating a counting rule or identifying and extending a numeric sequence; can work effectively with different two- and three-dimensional standard representations of objects or situations, for example devising a strategy to match one representation with another, or to compare different scenarios, or identify random experiment outcomes mathematically using standard conventions.
1	At this level students can recognise or modify and use an explicit simple model of a contextual situation. Students can choose between several such models to match the situation. For example, choose between and additive and a multiplicative model in a shopping context; choose among given two-dimensional objects to represent a familiar three-dimensional object; select one of several given graphs to represent growth of a population.

■ Figure 15.7 [Part 1/2] ■

Summary descriptions of the six proficiency levels on the mathematical process subscale *Employing mathematical concepts, facts, procedures and reasoning*

Level	What students can typically do
6	Students at or above Level 6 are typically able to employ a strong repertoire of knowledge and procedural skills in a wide range of mathematical areas. They can form and follow a multi-step strategy to solve a problem involving several stages; can apply reasoning in a connected way across several problem elements; can set up and solve an algebraic equation with more than one variable; can generate relevant data and information to explore problems, for example using a spreadsheet to sort and analyse data; are able to justify their results mathematically and to explain their conclusions and support them with well-formed mathematical arguments. At Level 6 students' work is consistently precise and accurate.
5	Students at Level 5 typically are able to employ a range of knowledge and skills to solve problems. They can sensibly link information in graphical and diagrammatic form to textual information. They can apply spatial and numeric reasoning skills to express and work with simple models in reasonably well-defined situations and where the constraints are clear. They usually work systematically, for example to explore combinatorial outcomes, and can typically sustain accuracy in their reasoning across a small number of steps and processes. They are generally able to work competently with expressions and can work with formulae and can use proportional reasoning; and are able to work with and transform data presented in a variety of forms.

■ Figure 15.7 [Part 2/2] ■

Summary descriptions of the six proficiency levels on the mathematical process subscale *Employing mathematical concepts, facts, procedures and reasoning*

Level	What students can typically do
4	At Level 4, students can typically identify relevant data and information from contextual material and use it to perform such tasks as calculating distances, and using proportional reasoning to apply a scale factor, convert different units to a common scale, or to relate different graph scales to each other. They are able to work flexibly with distance-time-speed relationships, and can carry out calculations in contexts that require a sequence of arithmetic calculations. They show some ability to use algebraic formulations, to follow a straightforward strategy and describe it.
3	Students at Level 3 frequently have sound spatial reasoning skills enabling them, for example, to use the symmetry properties of a figure, or to recognise patterns presented in graphical form, or to use angle facts, to solve a geometric problem. Students at this level can connect two different mathematical representations, such as data in a table and in a graph, or an algebraic expression with its graphical representation, enabling them for example to understand the effect of changing data in one representation on the other. They typically show some ability to handle percentages, fractions and decimal numbers and to work with proportional relationships.
2	Students at Level 2 are able to apply small reasoning steps to make direct use of given information to solve a problem, for example to implement a simple calculation model, or to identify a calculation error, or to analyse a distance-time relationship, or to analyse a simple spatial pattern; at this level students show an understanding of place value in decimal numbers and can use that understanding to compare numbers presented in a familiar context; can correctly substitute values into a simple formula; can recognise which of a set of given graphs correctly represents a set of percentages and can apply reasoning skills to understand and explore different kinds of graphical representations of data; and they typically show some insight into simple probability concepts.
1	Students at Level 1 can identify simple data relating to a real-world context, for example presented in a structured table or in an advertisement where the text and data labels match directly; can perform practical tasks such as decomposing money amounts into lower denominations; use direct reasoning from textual information that points to an obvious strategy to solve a given problem, particularly where the mathematical procedural knowledge required would be limited, for example, to arithmetic operations with whole numbers, or to ordering and comparing whole numbers; they demonstrate a partial understanding of graphing techniques and conventions; and can make use of symmetry properties to explore characteristics of a figure such as comparing side lengths and angles.

■ Figure 15.8 ■

Summary descriptions of the six proficiency levels on the mathematical process subscale *Interpreting, applying and evaluating mathematical outcomes*

Level	What students can typically do
6	At Level 6, students are able to link multiple complex mathematical representations in an analytic way to identify and extract data and information that enables contextual questions to be answered, and are able to effectively present their interpretations and conclusions in written form. For example they may interpret two time-series graphs in relation to different contextual conditions; or link a relationship expressed both in a graph and in numeric form (such as in a price calculator) or in a spreadsheet and graph, to present an argument or conclusion about contextual conditions. Students at this level are also typically able to apply mathematical reasoning to data or information presented in order to generate a chain of linked steps to support a conclusion (for example, analysis of a map using scale information; or analysis of a complex algebraic formula in relation to the variables represented; or translating data into a new time-frame; or performing a three-way currency conversion; or systematic use of a data generation tool to find the information needed to answer a question). Students at this level are able to bring together analysis and data and their interpretation across several different problem elements or across different questions about a context, showing a depth of insight and a capacity for sustained reasoning.
5	At Level 5, students are able to combine several processes in order to formulate conclusions based on interpretation of mathematical information with respect to context, such as formulating or modifying a model, solving an equation or carrying out computations, and using several reasoning steps to make the links to the identified context elements. At this level, students are able to make links between context and mathematics involving spatial or geometric concepts and complex statistical and algebraic concepts. They can easily interpret and evaluate a set of plausible mathematical representations, such as several graphs, to identify which one best reflects the contextual elements under analysis. Students at this level have begun to develop the ability to communicate conclusions and interpretations in written form.
4	At Level 4 students are typically able to apply appropriate reasoning steps, possibly multiple steps, to extract information from a complex mathematical situation, and to interpret complicated mathematical objects, including algebraic expressions. They can interpret complex graphical representations to identify data or information that answers a question; can perform a calculation or data manipulation (for example in a spreadsheet) to generate additional data needed to decide whether a constraint (such as a measurement condition, or a size comparison) is met; they can interpret simple statistical or probabilistic statements in such contexts as public transport, or health and medical test interpretation to link the meaning of the statements to the underlying contextual issues; they can conceptualise a change needed to a calculation procedure in response to a changed constraint; they can analyse two data samples, for example relating to a manufacturing process, to make comparisons and draw and express conclusions.
3	Students at Level 3 begin to show the ability to use reasoning, including spatial reasoning, to support their interpretations of mathematical information in order to make inferences about features of the context. They combine reasoning steps systematically to make various connections between mathematical and contextual material or when required to focus on different aspects of a context, for example where a graph shows two data series or a table contains data on two variables that must be actively related to each other to support a conclusion. They are able to test and explore alternative scenarios, using reasoning to interpret the possible effects of changing some of the variables under observation. They can use appropriate calculation steps to assist their analysis of data and to support the formation of conclusions and interpretations, including calculations involving proportions and proportional reasoning, and in situations where systematic analysis across several related cases is needed. At this level students can interpret and analyse relatively unfamiliar data presentations to support their conclusions.
2	At Level 2, students link contextual elements of the problem to the mathematics, for example by performing appropriate calculations or reading tables. Students at this level can typically make comparisons repeatedly across several similar cases; for example they can interpret a bar graph to identify and extract data to apply in a comparative condition where some insight is required. They can apply basic spatial skills to make connections between a situation presented visually and its mathematical elements; they can identify and carry out necessary calculations to support such comparisons as costs across several contexts; and they may be able to interpret a simple algebraic expression as it relates to a given context.
1	At Level 1, students are able to interpret data or information expressed in a direct way in order to answer questions about the context described. They can interpret given data to answer questions about simple quantitative relational ideas (such as 'larger', 'shorter time', 'in between') in a familiar context, for example by evaluating measurements of an object against given criterion values, or by comparing average journey times for two methods of transport, or comparing specified characteristics of a small number of similar objects. Similarly, they can make simple interpretations of data in a timetable or schedule to identify times or events. Students at this level may show rudimentary understanding of such concepts as randomness and data interpretation, for example by identifying the plausibility of a statement about chance outcomes of a lottery, or by understanding numeric and relational information in a well-labelled graph, and by understanding basic contextual implications of links between related graphs.

Figures 15.9, 15.10, 15.11 and 15.12 provide the summary descriptions of skills, knowledge and understanding required to complete tasks located within the defined bands for the mathematical content subscales: *Change and relationships, Space and Shape, Quantity* and *Uncertainty and data* respectively.

■ Figure 15.9 ■

Summary descriptions of the six proficiency levels on the mathematical content subscale *Change and relationships*

Level	What students can typically do
6	At Level 6, students use significant insight, abstract reasoning and argumentation skills and technical knowledge and conventions to solve problems involving relationships among variables and to generalise mathematical solutions to complex real-world problems. They are able to create and use an algebraic model of a functional relationship incorporating multiple quantities. They apply deep geometrical insight to work with complex patterns. And they are typically able to use complex proportional reasoning, and complex calculations with percentage to explore quantitative relationships and change.
5	At Level 5, students solve problems by using algebraic and other formal mathematical models, including in scientific contexts. They are typically able to use complex and multi-step problem-solving skills, and to reflect on and communicate reasoning and arguments, for example in evaluating and using a formula to predict the quantitative effect of change in one variable on another. They are able to use complex proportional reasoning, for example to work with rates, and they are generally able to work competently with formulae and with expressions including inequalities.
4	Students at Level 4 are typically able to understand and work with multiple representations, including algebraic models of real-world situations. They can reason about simple functional relationships between variables, going beyond individual data points to identifying simple underlying patterns. They typically employ some flexibility in interpretation and reasoning about functional relationships (for example in exploring distance-time-speed relationships) and are able to modify a functional model or graph to fit a specified change to the situation; and they are able to communicate the resulting explanations and arguments.
3	At Level 3, students can typically solve problems that involve working with information from two related representations (text, graph, table, formulae), requiring some interpretation, and using reasoning in familiar contexts. They show some ability to communicate their arguments. Students at this level can typically make a straightforward modification to a given functional model to fit a new situation; and they use a range of calculation procedures to solve problems, including ordering data, time difference calculations, substitution of values into a formula, or linear interpolation.
2	Students at Level 2 are typically able to locate relevant information on a relationship from data provided in a table or graph and make direct comparisons, for example to match given graphs to a specified change process. They can reason about the basic meaning of simple relationships expressed in text or numeric form by linking text with a single representation of a relationship (graph, table, simple formula), and can correctly substitute numbers into simple formulae, sometimes expressed in words. At this level, student can use interpretation and reasoning skills in a straightforward context involving linked quantities.
1	Students at Level 1 are typically able to evaluate single given statements about a relationship expressed clearly and directly in a formula, or in a graph. Their ability to reason about relationships, and to change in those relationships, is limited to simple expressions and to those located in familiar situations. They may apply simple calculations needed to solve problems related to clearly expressed relationships.

■ Figure 15.10 ■

Summary descriptions of the six proficiency levels on the mathematical content subscale *Space and shape*

Level	What students can typically do
6	At Level 6, students are able to solve complex problems involving multiple representations or calculations; identify, extract, and link relevant information, for example by extracting relevant dimensions from a diagram or map and using scale to calculate an area or distance; they use spatial reasoning, significant insight and reflection, for example by interpreting text and related contextual material to formulate a useful geometric model and applying it taking into account contextual constraints; they are able to recall and apply relevant procedural knowledge from their mathematical knowledge base such as in circle geometry, trigonometry, Pythagoras's rule, or area and volume formulae to solve problems; and they are typically able to generalise results and findings, communicate solutions and provide justifications and argumentation.
5	At Level 5, students are typically able to solve problems that require appropriate assumptions to be made, or that involve reasoning from assumptions provided and taking into account explicitly stated constraints, for example in exploring and analysing the layout of a room and the furniture it contains. They solve problems using theorems or procedural knowledge such as symmetry properties, or similar triangle properties or formulas including those for calculating area, perimeter or volume of familiar shapes; they use well-developed spatial reasoning, argument and insight to infer relevant conclusions and to interpret and link different representations, for example to identify a direction or location on a map from textual information.
4	Students at Level 4 typically solve problems by using basic mathematical knowledge such as angle and side-length relationships in triangles, and doing so in a way that involves multistep, visual and spatial reasoning, and argumentation in unfamiliar contexts; they are able to link and integrate different representations, for example to analyse the structure of a three dimensional object based on two different perspectives of it; and typically they can compare objects using geometric properties.
3	At Level 3, students are able to solve problems that involve elementary visual and spatial reasoning in familiar contexts, such as calculating a distance or a direction from a map or a GPS device; they are typically able to link different representations of familiar objects or to appreciate properties of objects under some simple specified transformation; and at this level students can devise simple strategies and apply basic properties of triangles and circles, and can use appropriate supporting calculation techniques such as scale conversions needed to analyse distances on a map.
2	At Level 2, students are typically able to solve problems involving a single familiar geometric representation (for example, a diagram or other graphic) by comprehending and drawing conclusions in relation to clearly presented basic geometric properties and associated constraints. They can also evaluate and compare spatial characteristics of familiar objects in a situation where given constraints apply (such as comparing the height or circumference of two cylinders having the same surface area; or deciding whether a given shape can be dissected to produce another specified shape).
1	Students at Level 1 can typically recognise and solve simple problems in a familiar context using pictures or drawings of familiar geometric objects and applying basic spatial skills such as recognising elementary symmetry properties, or comparing lengths or angle sizes, or using procedures such as dissection of shapes.

■ Figure 15.11 ■

Summary descriptions of the six proficiency levels on the mathematical content subscale *Quantity*

Level	What students can typically do
6	At Level 6 and above, students conceptualise and work with models of complex quantitative processes and relationships; devise strategies for solving problems; formulate conclusions, arguments and precise explanations; interpret and understand complex information, and link multiple complex information sources; interpret graphical information and apply reasoning to identify, model and apply a numeric pattern. They are able to analyse and evaluate interpretive statements based on data provided; work with formal and symbolic expressions; plan and implement sequential calculations in complex and unfamiliar contexts, including working with large numbers, for example to perform a sequence of currency conversions, entering values correctly and rounding results. Students at this level work accurately with decimal fractions; they use advanced reasoning concerning proportions, geometric representations of quantities, combinatorics and integer number relationships; and they interpret and understand formal expressions of relationships among numbers, including in a scientific context.
5	At Level 5, students are able to formulate comparison models and compare outcomes to determine best price; interpret complex information about real-world situations (including graphs, drawings and complex tables, for example two graphs using different scales); they are able to generate data for two variables and evaluate propositions about the relationship between them. Students are able to communicate reasoning and argument; recognise the significance of numbers to draw inferences; provide a written argument evaluating a proposition based on data provided. They can make an estimation using daily life knowledge; calculate relative and/or absolute change; calculate an average; calculate relative and/or absolute difference, including percentage difference, given raw difference data; and they can convert units (for example calculations involving areas in different units).
4	At Level 4, students are typically able to interpret complex instructions and situations; relate text-based numerical information to a graphic representation; identify and use quantitative information from multiple sources; deduce system rules from unfamiliar representations; formulate a simple numeric model; set up comparison models; and explain their results. They are typically able to carry out accurate and more complex or repeated calculations, such as adding 13 given times in hour/minute format; carry out time calculations using given data on distance and speed of a journey; perform simple division of large multiples in context; carry out calculations involving a sequence of steps and accurately apply a given numeric algorithm involving a number of steps. Students at this level can perform calculations involving proportional reasoning, divisibility or percentages in simple models of complex situations.
3	At Level 3, students typically use basic problem-solving processes, including devising a simple strategy to test scenarios, understand and work with given constraints, use trial and error, and use simple reasoning in familiar contexts. At this level students typically can interpret a text description of a sequential calculation process, and correctly implement the process; identify and extract data presented directly in textual explanations of unfamiliar data; interpret text and diagrams describing a simple pattern; perform calculations including working with large numbers, calculations with speed and time, conversion of units (for example from an annual rate to a daily rate). They understand place value involving mixed 2- and 3-decimal values and including working with prices; and are typically able to order a small series of (4) decimal values; calculate percentages of up to 3-digit numbers; and apply calculation rules given in natural language.
2	At Level 2, students can typically interpret simple tables to identify and extract relevant quantitative information; interpret a simple quantitative model (such as a proportional relationship) and apply it using basic arithmetic calculations. They are able to identify the links between relevant textual information and tabular data to solve word problems; interpret and apply simple models involving quantitative relationships; identify the simple calculation required to solve a straight-forward problem; carry out simple calculations involving the basic arithmetic operations, as well as ordering 2- and 3-digit whole numbers and decimal numbers with one or two decimal places, and calculate percentages.
1	At Level 1, students are typically able to solve basic problems in which relevant information is explicitly presented; the situation is straightforward and very limited in scope. Students at this level are able to handle situations where the required computational activity is obvious and the mathematical task is basic, such as a one-step simple arithmetic operation, or to total the columns of a simple table and compare the results; they can typically read and interpret a simple table of numbers; they can extract data and perform simple calculations; use a calculator to generate relevant data, extrapolate from the data generated, using reasoning and calculation with a simple linear model.

■ Figure 15.12 ■

Summary descriptions of the six proficiency levels on the mathematical content subscale *Uncertainty and data*

Level	What students can typically do
6	At Level 6, students are able to interpret, evaluate and critically reflect on a range of complex statistical or probabilistic data, information and situations to analyse problems. Students at this level bring insight and sustained reasoning across several problem elements; they understand the connections between data and the situations they represent and are able to make use of those connections to explore problem situations fully; they bring appropriate calculation techniques to bear to explore data or to solve probability problems; and they can produce and communicate conclusions, reasoning and explanations.
5	At Level 5, students are typically able to interpret and analyse a range of statistical or probabilistic data, information and situations to solve problems in complex contexts that require linking of different problem components. They can use proportional reasoning effectively to link sample data to the population they represent, can appropriately interpret data series over time and are systematic in their use and exploration of data. Students at this level can use statistical and probabilistic concepts and knowledge to reflect, draw inferences and produce and communicate results.
4	Students at Level 4 are typically able to activate and employ a range of data representations and statistical or probabilistic processes to interpret data, information and situations to solve problems. They can work effectively with constraints, such as statistical conditions that might apply in a sampling experiment, and they can interpret and actively translate between two related data representations (such as a graph and a data table). Students at this level can perform statistical and probabilistic reasoning to make contextual conclusions.
3	At Level 3, students are typically able to interpret and work with data and statistical information from a single representation that may include multiple data sources, such as a graph representing several variables, or from two simple related data representations such as a simple data table and graph. They are able to work with and interpret descriptive statistical, probabilistic concepts and conventions in contexts such as coin tossing or lotteries and make conclusions from data, such as calculating or using simple measures of centre and spread. Students at this level can perform basic statistical and probabilistic reasoning in simple contexts.
2	Students at Level 2 are typically able to identify, extract and comprehend statistical data presented in a simple and familiar form such as a simple table, a bar graph or pie chart; they can identify, understand and use basic descriptive statistical and probabilistic concepts in familiar contexts, such as tossing coins or rolling dice. At this level students can interpret data in simple representations, and apply suitable calculation procedures that connect given data to the problem context represented.
1	At Level 1, students can typically identify and read information presented in a small table or simple well-labelled graph to locate and extract specific data values while ignoring distracting information, and to recognise how these relate to the context. Students at this level can recognise and use basic concepts of randomness to identify misconceptions in familiar experimental contexts such as lottery outcomes.

Levels of proficiency in problem solving

The computer-based assessment of problem solving was the major innovative component of the PISA 2012 survey. Six proficiency levels were defined and described, and these are presented in Figure 15.13. The scale is further illustrated in Volume V of the *PISA 2012 Results* (OECD, 2014b). For a discussion of factors influencing item difficulty in these problem solving items, see Philpot et al. (forthcoming).

■ Figure 15.13 ■

Summary descriptions of the six proficiency levels on the problem solving scale

Level	Score range	What students can typically do
6	Equal to or higher than 683.1 points	At Level 6, students can develop complete, coherent mental models of diverse problem scenarios, enabling them to solve complex problems efficiently. They can explore a scenario in a highly strategic manner to understand all information pertaining to the problem. The information may be presented in different formats, requiring interpretation and integration of related parts. When confronted with very complex devices, such as home appliances that work in an unusual or unexpected manner, they quickly learn how to control the devices to achieve a goal in an optimal way. Level 6 problem-solvers can set up general hypotheses about a system and thoroughly test them. They can follow a premise through to a logical conclusion or recognise when there is not enough information available to reach one. In order to reach a solution, these highly proficient problem-solvers can create complex, flexible, multi-step plans that they continually monitor during execution. Where necessary, they modify their strategies, taking all constraints into account, both explicit and implicit.
5	618.2 to less than 683.1 points	At Level 5, students can systematically explore a complex problem scenario to gain an understanding of how relevant information is structured. When faced with unfamiliar, moderately complex devices, such as vending machines or home appliances, they respond quickly to feedback in order to control the device. In order to reach a solution, Level 5 problem-solvers think ahead to find the best strategy that addresses all the given constraints. They can immediately adjust their plans or backtrack when they detect unexpected difficulties or when they make mistakes that take them off course.
4	553.3 to less than 618.2 points	At Level 4, students can explore a moderately complex problem scenario in a focused way. They grasp the links among the components of the scenario that are required to solve the problem. They can control moderately complex digital devices, such as unfamiliar vending machines or home appliances, but they don't always do so efficiently. These students can plan a few steps ahead and monitor the progress of their plans. They are usually able to adjust these plans or reformulate a goal in light of feedback. They can systematically try out different possibilities and check whether multiple conditions have been satisfied. They can form an hypothesis about why a system is malfunctioning, and describe how to test it.
3	488.4 to less than 553.3 points	At Level 3, students can handle information presented in several different formats. They can explore a problem scenario and infer simple relationships among its components. They can control simple digital devices, but have trouble with more complex devices. Problem-solvers at Level 3 can fully deal with one condition, for example, by generating several solutions and checking to see whether these satisfy the condition. When there are multiple conditions or inter-related features, they can hold one variable constant to see the effect of change on the other variables. They can devise and execute tests to confirm or refute a given hypothesis. They understand the need to plan ahead and monitor progress, and are able to try a different option if necessary.
2	423.4 to less than 488.4 points	At Level 2, students can explore an unfamiliar problem scenario and understand a small part of it. They try, but only partially succeed, to understand and control digital devices with unfamiliar controls, such as home appliances and vending machines. Level 2 problem-solvers can test a simple hypothesis that is given to them and can solve a problem that has a single, specific constraint. They can plan and carry out one step at a time to achieve a sub-goal, and have some capacity to monitor overall progress towards a solution.
1	358.5 to less than 423.4 points	At Level 1, students can explore a problem scenario only in a limited way, but tend to do so only when they have encountered very similar situations before. Based on their observations of familiar scenarios, these students are able only to partially describe the behaviour of a simple, everyday device. In general, students at Level 1 can solve straightforward problems provided there is only a simple condition to be satisfied and there are only one or two steps to be performed to reach the goal. Level 1 students tend not to be able to plan ahead or set sub-goals.

Levels of financial literacy

For the optional PISA 2012 assessment of financial literacy, five proficiency levels were defined and described. The factors identified to explain the variance in item difficulty included familiarity of experience with (financial) products, life stage relevance, understanding and use of financial terms, understanding and application of financial products, reading demands, conceptual understanding of numeracy, application of numeracy skills, and capacity to make effective (financial) decisions.

The proficiency descriptions are presented in Figure 15.14, and these are further explained and illustrated in Volume VI of the *PISA 2012 Results* (OECD, 2014c).

■ Figure 15.14 [Part 1/2] ■

Summary descriptions of the five proficiency levels on the financial literacy scale

Level	Score range	What students can typically do
5	Equal to or higher than 624.6 points	Students apply their understanding of a wide range of financial terms and concepts to contexts that may only become relevant to their lives in the long term. They analyse complex financial products. They take into account features of financial documents that are significant but unstated or not immediately evident, such as transaction costs. They work with a high level of accuracy and solve non-routine financial problems. They describe the potential outcomes of financial decisions, showing an understanding of the wider financial landscape, such as income tax.
4	549.9 to less than 624.6 points	Students apply their understanding of less common financial concepts and terms to contexts that will be relevant to them as they move towards adulthood, such as bank account management and compound interest in saving products. They interpret and evaluate a range of detailed financial documents such as bank statements, and explain the functions of less commonly used financial products. They make financial decisions taking into account longer-term consequences such as the impact of loan repayment on cost. They solve routine problems in less common financial contexts.

■ Figure 15.14 [Part 2/2] ■

Summary descriptions of the five proficiency levels on the financial literacy scale

Level	Score range	What students can typically do
3	475.1 to less than 549.9 points	Students apply their understanding of commonly used financial concepts, terms and products to situations that are relevant to them. They begin to consider the consequences of financial decisions and they make simple financial plans in familiar contexts. They make straightforward interpretations of a range of financial documents. They apply a range of basic numerical operations, including calculating percentages. They choose the numerical operations needed to solve routine problems in relatively common financial literacy contexts, such as budget calculations.
2	400.3 to less than 475.1 points	Students begin to apply their knowledge of common financial products and commonly used financial terms and concepts. They use given information to make financial decisions in contexts that are immediately relevant to them. They recognise the value of a simple budget. They interpret prominent features of everyday financial documents. They apply single basic numerical operations, including division, to answer financial questions. They show an understanding of the relationships between different financial elements, such as the amount of use and the costs incurred.
1	Less than 400.3 points	Students identify common financial products and terms, and interpret information relating to basic financial concepts. They recognise the difference between needs and wants and they make simple decisions on everyday spending. They recognise the purpose of everyday financial documents and apply single and basic numerical operations (addition, subtraction or multiplication) in financial contexts that they are likely to have experienced personally.

References

Bourque, M. L. (2009), *A History of NAEP Achievement Levels: Issues, Implementation, and Impact 1989-2009*, National Assessment Governing Board, Washington D.C.

Beaton, A. E. and **N. L. Allen** (1992), "Interpreting scales through scale anchoring", *Journal of Educational and Behavioural Statistics,* No. 17(2), pp. 191-204.

Masters, G., J. Lokan, B. Doig, S.T. Khoo, J. Lindsey, L. Robinson and **S. Zammit** (1990), *Profiles of Learning: The Basic Skills Testing Program in New South Wales 1989,* Australian Council for Educational Research, Hawthorn, Victoria.

Mossenson, L., P. Hill, and **G. Masters** (1987), *Tests of Reading Comprehension [TORCH],* Australian Council for Educational Research, Hawthorn, Victoria.

OECD (2014a), *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

OECD (2014b), *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208070-en

OECD (2014c), *PISA 2012 Results: Students and Money: Financial Literacy Skills for the 21st Century (Volume VI),* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208094-en

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264190511-en

OECD (2011), *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI)*, PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264112995-en

OECD (2010), *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume 1)*, PISA, OECD Publishing, Paris.
http://dx.doi.org/ 10.1787/9789264091450-en

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World (Volume 1: Analysis)*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264040014-en

OECD (2004), *Learning for Tomorrow's World: First results from PISA 2003*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264006416-en

OECD (2002), *Reading for Change: Performance and Engagement Across Countries, Results from PISA 2000,* PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264099289-en

Philpot, R., D. Ramalingam, B. McCrae, and **J. Dossey** (forthcoming), "Factors that influence the difficulty of problem solving items", in Csapó, B. and J. Funke (eds.), *The Nature of Problem Solving,* OECD Publishing, Paris.

Wright, B. and **M. Stone,** (1979), *Best Test Design*, MESA Press, Chicago.

16

Scaling Procedures and Construct Validation of Context Questionnaire Data

OVERVIEW

The PISA 2012 context questionnaires included numerous items on student characteristics, student family background, student perceptions, school characteristics and perceptions of school principals. While student and school context questionnaires were mandatory in all countries, 11 countries also administered the optional questionnaire to parents of the tested students. In addition, students were administered the Information and Communication Technology Questionnaire (ICT) in 42 countries and the Educational Career Questionnaire in 23 countries.

Some of the items were designed to be used in analyses as single items (for example, gender). However, most questionnaire items were designed to be combined in some way in order to measure latent constructs that cannot be observed directly. To these items, transformations or scaling procedures were applied to construct meaningful indices.

This chapter describes, for all five questionnaires, how indices based on one or more items were constructed and validated. As in previous PISA surveys, two different kinds of indices can be distinguished:

- simple indices: these indices were constructed through the arithmetical transformation or recoding of one or more items; and
- scale indices: these indices were constructed through the scaling of items. Typically, scale scores for these indices were estimates of latent traits derived through Item Response Theory (IRT) scaling of dichotomous or Likert-type items.

This chapter *(i)* outlines how simple indices were constructed, *(ii)* describes the methodology used for construct validation and scaling, *(iii)* details the construction and validation of scaled indices and *(iv)* illustrates the computation of the *PISA index of economic, social and cultural status (ESCS)*. Some indices have been used in previous PISA surveys and were constructed based on a similar scaling methodology (see Schulz, 2003; OECD, 2005). Other indices were based on the elaboration of the PISA 2012 questionnaire framework (see OECD, 2013, Chapter 6) and related to mathematics as the major domain of the fifth PISA cycle (see Chapter 1).

SIMPLE QUESTIONNAIRE INDICES

Student age

The age of a student (*AGE*) was calculated as the difference between the year and month of the testing and the year and month of a student's birth. Data on student's age were obtained from both the questionnaire and the student tracking forms. If the month of testing was not known for a particular student, the median month of testing for that country was used in the calculation. The formula for computing *AGE* was

16.1

$$AGE = \left(100 + T_y - S_y\right) + \frac{\left(T_m - S_m\right)}{12}$$

where T_y and S_y are the year of the test and the year of the students' birth of the tested student, respectively in two-digit format (for example "06" or "92"), and T_m and S_m are the month of the test and month of the students' birth, respectively. The result is rounded to two decimal places.

Study programme indices

PISA 2012 collected data on study programmes available to 15-year-old students in each country. This information was obtained through the student tracking form and the Student Questionnaire. In the final database, all national programmes are included in a separate variable (*PROGN*) where the first six digits represent the National Centre code, and the last two digits are the nationally specific programme code. All study programmes were classified using the International Standard Classification of Education (ISCED) (OECD, 1999). The following indices were derived from the data on study programmes: programme level (*ISCEDL*) indicating whether students were at the lower or upper secondary level (ISCED 2 or ISCED 3); programme designation (*ISCEDD*) indicating the designation of the study programme (A = general programmes designed to give access to the next programme level, B = programmes designed to give access to vocational studies at the next programme level, C = programmes designed to give direct access to the labour market, M = modular programmes that combine any or all of these characteristics); and programme orientation (*ISCEDO*) indicating whether the programme's curricular content was general, pre-vocational or vocational.

Highest occupational status of parents

Occupational data for both the student's father and student's mother were obtained by asking open-ended questions. The responses were coded to four-digit ISCO codes (ILO, 2007) and then mapped to the international socio-economic index of occupational status (*ISEI*) (Ganzeboom, 2010). In PISA 2012, the new ISCO and ISEI in their 2008 version were used rather than the 1988 versions that had been applied in the previous four cycles. For details regarding the update and results of analyses of the possible impact of this update, please see Chapter 17 of this report.

Three indices were calculated based on this information: father's occupational status (*BFMJ2*); mother's occupational status (*BMMJ1*); and the highest occupational status of parents (*HISEI*) which corresponds to the higher *ISEI* score of either parent or to the only available parent's *ISEI* score. For all three indices, higher *ISEI* scores indicate higher levels of occupational status.

Educational level of parents

Parental education is a second family background variable that is often used in the analysis of educational outcomes. Theoretically, it has been argued that parental education is more relevant for students' academic outcomes than parental occupation. Like occupation, the collection of internationally comparable data on parental education poses significant challenges, and less work has been done on internationally comparable measures of parental education than has been done on occupational status. The main difficulties with parental education relate to international comparability (education systems differ widely between countries and within countries over time), response validity (students are often unable to accurately report their parents level of education) and – especially with increasing migration – difficulties in the national mapping of parental qualifications gained elsewhere.

Students' responses regarding parental education were classified using ISCED (OECD, 1999). Indices on parental education were constructed by recoding educational qualifications into the following categories: (0) None, (1) ISCED 1 (primary education), (2) ISCED 2 (lower secondary), (3) ISCED Level 3B or 3C (vocational/pre-vocational upper secondary), (4) ISCED 3A (general upper secondary) and/or ISCED 4 (non-tertiary post-secondary), (5) ISCED 5B (vocational tertiary) and (6) ISCED 5A, 6 (theoretically oriented tertiary and post-graduate). Indices with these categories were provided for the students' mother (*MISCED*) and the students' father (*FISCED*). In addition, the *index of highest educational level of parents* (*HISCED*) corresponds to the higher ISCED level of either parent.

The *index of highest educational level of parents* was also recoded into estimated number of years of schooling (*PARED*). A mapping of ISCED levels of years of schooling is provided in Annex D.

Immigration background

Information on the country of birth of the students and their parents was also collected. Included in the database were three country-specific variables relating to the country of birth of the student, their mother and father (*COBN_S*, *COBN_M*, and *COBN_F*). The items ST20Q01, ST20Q02 and ST20Q03 were also recoded for the database into the following categories: (1) country of birth is same as country of assessment and (2) otherwise.

The *index of immigrant background* (*IMMIG*) was calculated from these variables with the following categories: (1) native students (those students who had at least one parent born in the country), (2) second generation students (those born in the country of assessment but whose parent(s) were born in another country) and (3) first-generation students (those students born outside the country of assessment and whose parents were also born in another country). Students with missing responses for either the student or for both parents were assigned missing values for this variable.

Language spoken at home

Students also indicated what language they usually spoke at home (*ST25*) and the database includes a variable (*LANGN*) containing a country-specific code for each language. In addition, an internationally comparable variable *ST25Q01* was derived from this information with the following categories: (1) language at home is the same as the language of assessment for that student and (2) language at home is another language.

Family structure

Information collected from students regarding their family structure formed the basis for the index *FAMSTRUC* with the following categories: "1" if "single parent family" (students living with only one of the following: mother, father, male guardian, female guardian), "2" if "two parent family" (students living with a father or step/foster father and a mother or step/foster mother) and "3" for students who do not live with their parents.

Relative grade

The relative grade index (*GRADE*) was computed to capture between country variation. It indicates whether students are at a modal grade in a country (value of 0) or whether they are below or above the modal grade (+x grades, -x grades).

Grade repetition

The grade repetition variable (*REPEAT*) was computed by recoding variables *ST07Q01*, *ST07Q02*, *ST07Q03*. It took the value of "1" if the student had repeated a grade in at least one level and the value of "0" if "No, never" was chosen at least one time, given that none of the repeated grade categories were chosen. The index is assigned a missing value if none of the three categories were ticked in any of three levels.

Out-of-school study time

Students were asked in open-ended format how much time they spent studying outside school, for example, with a tutor or parent (*ST57Q01- ST57Q06*). The index *OUTHOURS* was computed by summing the time spent studying for school subjects.

Learning time

Learning time in test language (*LMINS*) was computed by multiplying the number of minutes on average in the test language class by number of test language class periods per week (*ST69 and ST70*). Comparable indices were computed for mathematics (*MMINS*) and science (*SMINS*).

School Questionnaire indices

School size

The *index of school size* (*SCHSIZE*) contains the total enrolment at school based on the enrolment data provided by the school principal, summing the number of girls and boys at a school. This index was calculated in 2012 and in all previous cycles.

Proportion of girls enrolled at school

The *index on the proportion of girls at school* (*PCGIRLS*) is based on the enrolment data provided by the school principal (SC07), dividing the number of girls by the total number of girls and boys at a school. Prior to 2012, this index was also calculated in 2000 and 2006.

School type

Schools are classified as either public or private according to whether a private entity or a public agency has the ultimate power to make decisions concerning its affairs. As in previous PISA surveys, the index on school type (*SCHLTYPE*) has three categories, based on two questions: (1) government-independent private schools controlled by a non-government organisation or with a governing board not selected by a government agency which receive less than 50% of their core funding from government agencies, (2) government-dependent private schools controlled by a non-government organisation or with a governing board not selected by a government agency which receive more than 50% of their core funding from government agencies, (3) public schools controlled and managed by a public education authority or agency. This index was calculated in 2012 and in all previous cycles. In 2009 the variable name was *SCHTYPE*.

Availability of computers

School principals were asked to report the number of computers available at school. The *index of availability of computers* (*RATCMP15*) was the ratio of computers available to 15-year-olds for educational purposes to the total number of students in the modal grade for 15-year-olds. This was a new index in 2012.

As in previous cycles, in PISA 2012, the index *COMPWEB* was calculated as the ratio of number of computers available to 15-year-olds for educational purposes to the number of these computers that were connected to the web.

To obtain information on the educational use of technology, a question in the school questionnaire (SC13) asked about the proportion of time that the school expected students to use Internet/World Wide Web for homework, during lessons and for assignments or projects. No index was created, however.

Quantity of teaching staff at school

Principals were asked to report the number of full-time and part-time teachers at their school. Teachers in general and mathematics teachers were reported separately. However, since PISA 2006, the number of items about teachers in general has been reduced to capture only teachers in total, certified teachers, and teachers with an ISCED 5A qualification, rather than providing further break-downs by subject or specialist areas. For all of the following indices the number of part-time teachers contributed 0.5 and the number of full-time teachers 1.0 to the estimated numbers of teachers at school.

The student-teacher ratio (*STRATIO*) was obtained by dividing the number of enrolled students (index *SCHSIZE*) by the total number of teachers.

The proportion of fully certified teachers (*PROPCERT*) was computed by dividing the number of fully certified teachers by the total number of teachers.

The proportion of teachers with an ISCED 5A qualification (*PROPQUAL*) was calculated by dividing the number of these teachers by the total number of teachers.

The student-mathematics teacher ratio (SMRATIO) was obtained by dividing the number of enrolled students (index SCHSIZE) by the total number of mathematics teachers.

The proportion of mathematics teachers (PROPMATH) was computed by dividing the number of mathematics teachers by the total number of teachers.

The proportion of mathematics teachers with an ISCED 5A qualification (PROPMA5A) was calculated by dividing the number of these teachers by the total number of mathematics teachers.

Use of assessments

School principals were asked to indicate whether or not assessments of 15-year-old students were used for the following purposes at school: *i)* informing parents about progress; *ii)* for decisions about students' retention or promotion; *iii)* grouping students for instructional purposes; *iv)* comparison with district or national performance; *v)* monitoring the school's yearly progress; *vi)* judgements about teachers' effectiveness; *vii)* identification of areas for improvement; and *viii)* comparison with other schools. The index *use of assessments* (*ASSESS*) was calculated as the sum of "yes" responses to these eight items. Although the variable name of this index is the same as in some of the previous cycles, this index is not comparable with those cycles.

Class size

The average class size (CLSIZE) was derived from one of nine possible categories, ranging from "15 students or fewer" to "More than 50 students" for the average class size of the test language in the sampled schools. The midpoint of each response category was used for CLSIZE, resulting in a value of 13 for the lowest category, and a value of 53 for the highest.

Extra-curricular activities at school

School principals were asked to report what extra-curricular activities their schools offered to 15-year-old students (SC16).

The *index of creative extra-curricular activities at school* (CREACTIV) was computed as the total number of the following activities that occurred at school: *i)* band, orchestra or choir; *ii)* school play or school musical; and *iii)* art club or art activities.

Mathematics activities at school

The index of mathematics-related extra-curricular activities at school (*MACTIV*) was computed as follows. First, the question SCQ21 was assigned the value of '1' if the purpose of additional mathematics lessons was "enrichment mathematics only", "remedial mathematics only", or "without differentiation depending on the prior achievement level of the students" whereas it was assigned the value of '2' if "both enrichment mathematics and remedial mathematics" was the reported purpose. Second, each of three items about a mathematics club (*SC16Q05*), mathematics competitions (*SC16Q06*), or club with a focus on computers/ ICT (*SC16Q08*) was assigned the value of '1' if a school reported to offer these activities to 15-year-old students. Where a school did not offer one of these three activities, the corresponding

variable received the value of '0'. Third, these recoded variables were summed up to result in a range of "0" to "5" for *MACTIV*. For example, if the purpose of additional lessons was "both enrichment mathematics and remedial mathematics" and the school offered a mathematics club, but not an ICT club or mathematics competitions, the value of *MACTIV* was coded as "3".

The index of mathematics extension courses offered at school (*MATHEXC*) was created by assigning schools to one of three different categories based on the mathematics extension course types offered at school. Schools that offered additional mathematics courses without differentiation based on prior achievement were assigned a '1', schools that offered either enrichment mathematics only or remedial mathematics classes only were assigned a '2' and schools that offered both enrichment and remedial mathematics classes were assigned a '3'.

School selectivity

As in previous cycles, school principals were asked about admittance policies at their school, including placement tests and recommendation by feeder schools. The response scale was modified in 2012 from indicating whether or not a policy was a prerequisite to a frequency scale of "never", "sometimes", and "always".

In 2012, an index of academic school selectivity (*SCHSEL*) was computed by assigning schools to one of three categories based on how often two factors, namely student academic performance and recommendation of feeder schools were considered when admitting students to the school as follows: (1) the two factors were never considered, (2) at least one factor was considered sometimes but neither always and (3) at least one factor was considered always.

Ability grouping for mathematics classes

School principals were asked to report the extent to which their mathematics instruction catered for students with different abilities (SC15). The first two items asked about the occurrence of ability grouping into different classes either with similar content but different difficulty levels or with different content, one about ability grouping within classes and one about the application of different pedagogies within a class rather than ability grouping. Response categories were "For all classes", "For some classes" and "Not for any classes".

An index of ability grouping between mathematics classes (*ABGMATH*) was derived from the first two items by assigning schools to three categories: (1) schools with no ability grouping for any classes, (2) schools with one of these forms of ability grouping between some classes and (3) schools with one of these forms of ability grouping for all classes.

School responsibility for resource allocation

An index of the relative level of responsibility of school staff in allocating resources (*RESPRES*) was derived from six items of the school principals' report regarding who had considerable responsibility for tasks related to resource allocation ("Selecting teachers for hire", "Firing teachers", "Establishing teachers' starting salaries", "Determining teachers' salaries increases", "Formulating the school budget", "Deciding on budget allocations within the school"). The index was calculated on the basis of the ratio of "yes" responses for school governing board, principal or teachers to "yes" responses for regional/local education authority or national educational authority. Higher values on the scale indicated relatively higher levels of school responsibility in this area. The index was standardised to having an OECD mean of '0' and a standard deviation of '1' for the pooled data set with equally weighted country samples. This index was also created in 2006 and 2009 PISA cycles.

School responsibility for curriculum and assessment

An index of the relative level of responsibility of school staff in issues relating to curriculum and assessment (*RESPCUR*) was computed from the school principal's report regarding who had responsibility for four aspects of curriculum and assessment, namely "Establishing student assessment policies", "Choosing which textbooks are used", "Determining course content", and "Deciding which courses are offered". The index was calculated on the basis of the ratio of "yes" responses for school governing board, principal or teachers on the one hand to "yes" responses for regional/local education authority or national educational authority on the other hand. Higher values indicated relatively higher levels of school responsibility in this area. The index was standardised to having an OECD mean of '0' and a standard deviation of '1' for the pooled data with equally weighted country samples). This index was also created in all previous PISA cycles, although in PISA 2009 the variable name was *RESPCURR*.

Parent Questionnaire indices

Educational level of parents

Administration of the parent questionnaire in PISA 2012 provided the opportunity to collect data on parental education directly from the parents in addition to the data provided from their children in the student questionnaire. Similar to the student questionnaire data, parental educational levels were classified using ISCED (OECD, 1999). The question format differed from the one used in the student questionnaire as only four items were included with the dichotomous response categories of “yes” or “no’’.

Indices were constructed by taking the highest level for father and mother and having the following categories: (0) None, (1) ISCED 3A (general upper secondary) and/or ISCED 4 (non-tertiary post-secondary), (2) ISCED 5B (vocational tertiary), (3) ISCED 5A, 6 (theoretically oriented tertiary and post-graduate). Indices with these categories were computed for mother (*PQMISCED*) and father (*PQFISCED*). Highest educational level of parents (*PQHISCED*) corresponds to the higher ISCED level of either parent.

Occupational status of parents reported by parents

Occupational data for both the student’s father and student’s mother were obtained by asking open-ended questions of the parents themselves. The responses were coded into four-digit ISCO-08 codes and then mapped to the international socio-economic index of occupational status (*ISEI*) (Ganzeboom et al., 2010). Three indices were obtained from these data: father’s occupational status (*PQBFMJ*); mother’s occupational status (*PQBMMJ*); and the highest occupational status of parents (*PQHISEI*) which corresponds to the higher *ISEI* of either parent or to the only available parent’s *ISEI* score. For all three indices, higher *ISEI* indicates higher levels of occupational status.

Occupational aspirations of parents for their child

In addition, parents were asked about the occupational aspirations for their child. The response was coded into a four-digit ISCO-08 code and then mapped to the international socio-economic index of occupational status (Ganzeboom et al., 2010). This resulted in the *PQOCCASP* index.

Immigration status of parents

Two indices reflecting the immigrant status of parents were calculated based on responses to the countries of birth of the students’ parents as well as the students’ grandparents: father’s immigration status (*PQIMMIGF*) and mother’s immigration status (*PQIMMIGM*). The indices were coded to have the following categories: (1) native parent (those parents who had at least one parent (students’ grandparent) born in the country of the assessment); (2) second generation parent (those born in the country of assessment but whose parent(s) were born in another country); (3) first-generation parent (those parent born outside the country of assessment and whose parents were also born outside the country of assessment). Parents with missing responses for either the parent or for both grandparents were assigned missing values for these indices.

Citizenship of parents

Two indices regarding the citizenship of parents were calculated: father’s citizenship (*PQCITIZF*) and mother’s citizenship (*PQCITIZM*). The indices were coded to have the following categories: (1) parent has citizenship of the country of the assessment only; (2) parent has citizenship of the country of the assessment and another country (or countries); (3) parent has citizenship of a country (or countries) other than the country of the assessment.

Language spoken by parents at home

Two indices reflecting the language spoken by parents at home were calculated. *PQLANGNM* is a country-specific three digit code for the language the mother spoke at home most of the time. Responses to this question were transformed into an internationally comparable variable, *PA25Q01*, with the following categories: (1) mother’s language at home was same as the language in which the student took the test; (2) mother’s language at home was another language.

PQLANGNF is a country-specific three digit code for the language the father spoke at home most of the time. Responses to this question were transformed into an internationally comparable variable, *PA25Q02*, with the following categories: (1) father’s language at home was same as the language in which the student took the test; (2) father’s language at home was another language.

SCALING METHODOLOGY AND CONSTRUCT VALIDATION

Scaling procedures

Most questionnaire items were scaled using IRT scaling methodology. With the One-Parameter (Rasch) model (Rasch, 1960) for dichotomous items, the probability of selecting category 1 instead of 0 is modelled as

16.2

$$P_i(\theta_n) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

where $P_i(\theta_n)$ is the probability of person *n* to score 1 on item *i*. θ_n is the estimated latent trait of person *n* and δ_i the estimated location of item *i* on this dimension. For each item, item responses are modelled as a function of the latent trait θ_n.

In the case of items with more than two (*k*) categories (as for example with Likert-type items) this model can be generalised to the Partial credit model (Masters and Wright, 1997), which takes the form of

16.3

$$P_{x_i}(\theta_n) = \frac{\exp \sum_{k=0}^{x} (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^{h} (\theta_n - \delta_i + \tau_{ik})} \qquad x_i = 0, 1, \ldots, m_i$$

where $P_{xi}(\theta_n)$ denotes the probability of person *n* to score *x* on item *i* out of the m_i possible scores on the item. θ_n denotes the person's latent trait, the item parameter δ_i gives the location of the item on the latent continuum and τ_{ij} denotes an additional step parameter.[1]

International item parameters were obtained using the *ConQuest* software (Adams, Wu, and Wilson, 2012a). The calibration samples consisted of randomly selected sub-samples:

- For the calibration of student item parameters except items involved in household possessions and home background indices, sub-samples of 750 students were randomly selected within each country sample available at the time of calibration. As final student weights had not been available at the time the calibration sample was drawn, the random selection was based on preliminary student weights obtained from the ratio between sampled and enrolled students within explicit sampling strata. The final calibration sample included data from 48 000 students.
- For the calibration of parent item parameters, all available data from the parent questionnaires were merged with the student calibration sample described above.
- For the calibration of student item parameters for the items involved in household possessions and home background indices the calibration sample was drawn from all cycles for trend purposes. For the first four cycles, 500 cases were drawn from each participating country. For the fifth cycle, PISA 2012, 750 cases were drawn from each country to emphasise the PISA 2012 data as the basis for trend. The final calibration sample included data from 154 541 students.
- For the calibration of school item parameters, all available data were used and countries were weighted equally.

Once the international item parameters had been estimated from the calibration sample, Weighted Likelihood Estimate (WLE; Warm, 1989) was used to obtain individual participant scores. The WLEs were derived using the *ConQuest* software (Wu, Adams and Wilson, 2012a) with pre-calibrated item parameters.

WLEs were transformed to an international metric with an OECD average of zero and an OECD standard deviation of one. The transformation was achieved by applying the formula

16.4

$$\theta_n' = \frac{\theta_n - \bar{\theta}_{OECD}}{\sigma_{\theta(OECD)}}$$

where θ_n' are the scores in the international metric, θ_n the original WLE in logits, and $\bar{\theta}_{\text{OECD}}$ is the OECD mean of logit scores with equally weighted country samples. $\sigma_{\theta(\text{OECD})}$ is the corresponding OECD standard deviation of the original WLEs. Means and standard deviations (S.D.) used for the transformation into the international metric are shown in Table 16.1.

Table 16.1 **OECD means and standard deviations of Weighted Likelihood Estimates (WLEs)**

Student-level indices	N	Mean	S.D.
ANXMAT	22008	-0.27	1.87
ATSCHL	21939	1.01	1.45
ATTLNACT	21908	2.71	1.97
BELONG	21980	1.36	1.57
CLSMAN	21900	0.97	1.45
COGACT	21989	0.35	1.26
CULTDIST	1161	1.13	1.90
CULTPOS	33062	0.17	1.41
DISCLIMA	21989	0.97	2.01
ENTUSE	27412	-0.02	0.77
EXAPPLM	21940	0.00	1.21
EXPUREM	21900	2.39	2.23
FAILMAT	21994	0.03	0.96
FAMCON	21284	0.15	0.69
FAMCONC	21201	0.86	0.84
HEDRES	33400	2.06	1.39
HERITCUL	1194	2.00	2.31
HOMEPOS	33495	1.45	0.93
HOMSCH	27234	-0.92	1.18
HOSTCUL	1204	2.53	2.50
ICTATTNEG	26902	-0.57	1.74
ICTATTPOS	26978	2.16	2.09
ICTHOME	27630	0.59	0.76
ICTRES	33439	1.34	1.38
ICTSCH	27614	-0.21	1.15
INFOCAR	14374	-0.35	1.34
INFOJOB1	7019	-1.23	1.15
INFOJOB2	7019	-0.88	1.26
INSTMOT	22101	1.32	2.78
INTMAT	22119	-0.82	2.93
LMINS	20053	214.69	93.93
MATBEH	21993	-1.55	1.12
MATHEFF	22080	1.15	1.50
MATINTFC	20960	0.06	1.68
MATWKETH	22003	0.75	1.77
MTSUP	21954	1.33	1.84
OPENPS	21962	0.85	1.50
PERSEV	21984	0.28	0.91
SCMAT	22008	-0.13	2.36
STUDREL	22004	1.32	2.01
SUBNORM	22104	0.35	1.17
TCHBEHFA	22000	-0.28	1.35
TCHBEHSO	22019	-0.98	1.06
TCHBEHTD	22041	0.54	1.14
TEACHSUP	22070	1.03	1.73
USEMATH	26921	-1.57	1.57
USESCH	27152	-1.60	1.22
WEALTH	33477	1.25	1.10
ANCATSCHL	21527	0.07	1.55
ANCATTLNACT	21522	1.39	3.08
ANCBELONG	21570	0.21	1.63
ANCCLSMAN	21607	0.00	1.58
ANCCOGACT	21593	-0.36	1.47
ANCINSTMOT	10852	-0.18	2.41
ANCINTMAT	10852	-1.91	3.52
ANCMATWKETH	10800	-0.30	2.18
ANCMTSUP	21619	0.17	1.90
ANCSCMAT	21546	-0.84	2.33
ANCSTUDREL	21590	-0.01	1.99
ANCSUBNORM	10859	-0.39	1.52

Parent-level indices	N	Mean	S.D.
PARINVOL	7832	-2.04	1.56
PARSUPP	7902	0.59	1.02
PQMCAR	7873	-0.44	1.68
PQMIMP	7880	2.99	2.77
PQSCHOOL	7882	1.83	2.09

School-level indices	N	Mean	S.D.
SCMATBUI	31458	1.06	1.61
SCMATEDU	32427	1.06	1.45
STUDCLIM	32338	1.34	1.60
TCFOCST	32139	1.46	1.70
TCHPARTI	32649	-1.99	1.75
TCMORALE	32413	3.61	2.43
TCSHORT	32346	-2.32	1.85
TEACCLIM	31676	1.49	1.30
SCHAUTON	32649	1.44	1.88
LEADINST	32082	0.42	1.23
LEADTCH	31779	0.60	1.33
LEADPD	31871	1.22	1.46
LEADCOM	32131	0.27	1.13

Construct validation

The development of comparable measures of student background, attitudes and perceptions is a major goal of PISA. Cross-country validity of these constructs is of particular importance as measures derived from questionnaires are often used to predict differences in student performance within and across countries and are, thus, potential sources of policy-relevant information about ways of improving educational systems. There are different methodological approaches for validating questionnaire constructs, each with their advantages and limitations.

Cronbach's alpha was used to check internal consistency of each scaled index within the countries and to compare it between the countries. For some indices, some countries opted to delete one or two questions. Strictly speaking, this constituted a different index and, therefore, a footnote was added in the tables to note which item had been deleted.

Cross-country validity of the constructs not only requires a thorough and closely monitored process of translation into different languages. It also makes assumptions about having measured similar characteristics, attitudes and perceptions in different national and cultural contexts. Psychometric techniques can be used to analyse the extent to which constructs have consistent construct validity across participating countries. This is done by first checking the reliability of the scales across individual countries and then correlations are also estimated for certain scales which are thought to be related. These correlations should be consistent across countries.

Patterns of consistent relationships between certain indices across countries can be seen, for example, in Table 16.20 with similar correlations across the OECD countries between the indices, *teacher behaviour - student orientation (TCHBEHSO) and teacher behaviour - formative assessment (TCHBEHFA)*. Table 16.21 shows that the correlations between these two indices are also similar for the partner countries and economies. Similar results are found in Table 16.20 for OECD countries and in Table 16.21 for partner countries and economies for the correlations between the indices, *classroom management (CLSMAN)* and *disciplinary climate* (*DISCLIMA*). Likewise, Table 16.39 for OECD countries and Table 16.40 for partner countries and economies show the expected positive relationship between the indices, *attitude towards school - learning activities (ATTLNACT)* and *attitude towards school - learning outcomes (ATSCHL)* across countries.

Describing questionnaire scale indices

As in previous PISA surveys, in PISA 2012 categorical items from the context questionnaires were scaled using IRT modelling. WLEs (logits) for the latent dimensions were transformed to scales with an OECD average of 0 and a standard deviation of 1 (with equally weighted samples). It is possible to interpret these scores by comparing individual scores or group average scores to the OECD mean, but the individual scores do not reveal anything about the actual item responses and it is impossible to determine from scale score values to what extent respondents endorsed the items used for the measurement of the latent variable. However, the scaling model used to derive individual scores allows descriptions of these scales by mapping scale scores to (expected) item responses.[2]

Item characteristics can be described using the parameters of the partial credit model by summing for each category its probability of being chosen with the probabilities of all higher categories. This is equivalent to computing the odds of scoring higher than a particular category.

The results of plotting these cumulative probabilities against scale scores for a fictitious item are displayed in Figure 16.1. The three vertical lines denote those points on the latent continuum where it becomes more likely to score >0, >1 or >2. These locations, Γ_k , are Thurstonian thresholds that can be obtained through an iterative procedure that calculates summed probabilities for each category at each (decimal) point on the latent variable.

Summed probabilities are not identical with expected item scores and have to be understood in terms of the probability to score *at least* a particular category. Other ways of describing the item characteristics based on the partial credit model are item characteristic curves (by plotting the individual category probabilities) and expected item score curves (for a more detailed description see Masters and Wright, 1997).

Thurstonian thresholds can be used to indicate those points on a scale for each item category, at which respondents have a 0.5 probability to score this category or higher. For example, in the case of Likert-type items with categories "Strongly disagree" (SD), "Disagree" (D), "Agree" (A) and "Strongly agree" (SA) it is possible to determine at what point of a scale a respondent has a 50% chance to agree with the item.

■ Figure 16.1 ■

Summed category probabilities for fictitious items

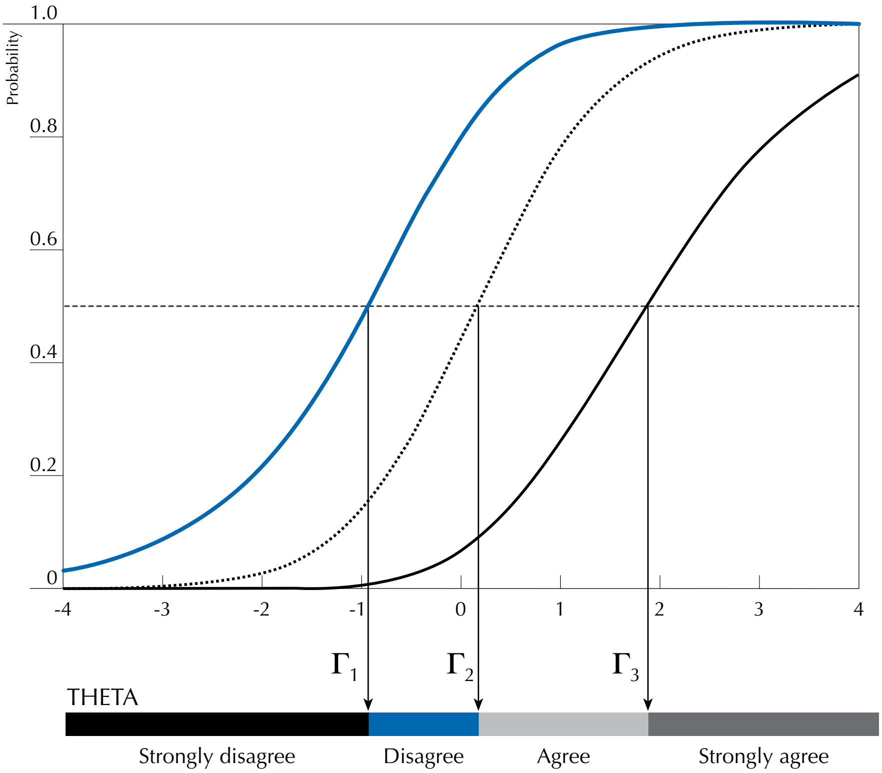
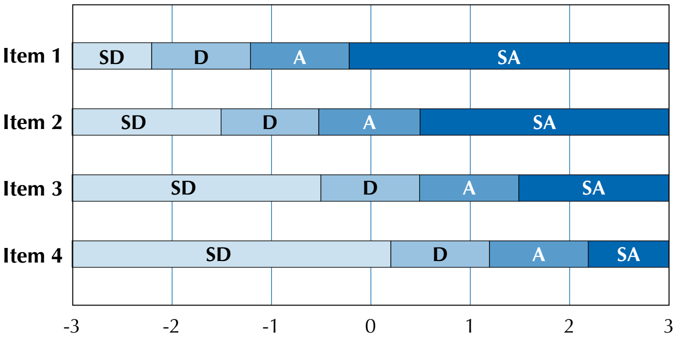


Figure 16.2 shows the fictitious example of an item map for a fictitious scale that consists of four Likert-type items, each with four response options, namely "Strongly disagree" (SD), "Disagree" (D), "Agree" (A) and "Strongly agree" (SA). Interpretation of this item map is as follows:

- Students with a score of –2 (that is, 2 standard deviations below the OECD average) have a 0.5 probability to disagree, agree or strongly agree (or not to disagree strongly with item 1), but they have more than a 50% chance to strongly disagree with the other three items.
- Students with a score of –1 (one standard deviation below the OECD average), have already a probability greater than 0.5 to agree with the first item, but they would still be expected to disagree with item 2 or even to strongly disagree with item 3 and 4.

■ Figure 16.2 ■

Fictitious example of item map



- Likewise, students with a score of 1 (one standard deviation above the OECD average) would have a probability greater than 0.5 to strongly agree with the first two items, but still have a probability lower than 0.5 to agree with item 4.

Item maps can help to illustrate the relationship between scores and item responses. For example, even scores of one standard deviation below the OECD average on an attitudinal scale could still indicate affirmative responses. This would not be revealed by the international metric, which have to be interpreted relative to the OECD average, but can be concluded from the corresponding item map.

QUESTIONNAIRE SCALE INDICES

Student scale indices

Household possessions

Collecting information about household possessions as indicators of family wealth has received much attention in international studies in the field of education (Spiezia, 2010; Traynor and Raykov, 2013). Household assets are believed to capture wealth better than income because they reflect a more stable source of wealth.

In PISA 2012, students reported the availability of 14 household items at home (ST26). In addition, countries added three specific household items that were seen as appropriate measures of family wealth within the country's context.

Four indices were derived from these items: *i) family wealth possessions* (*WEALTH*), *ii) cultural possessions* (*CULTPOS*), *iii) home educational resources* (*HEDRES*), and *iv) home possessions* (*HOMEPOS*). The last index was a summary index of all household items from the units ST26 and ST27 and also included the variable indicating the number of books at home (ST28) which was recoded from the original six categories into three: (0) 0-25 books, (1) 26-100 books, (2) more than 100. Questions ST27Q01 and ST27Q02 were recoded from the original four categories into three: (0) "None or one", (1) "Two" (2) "Three or more". Questions ST27Q03, ST27Q04 and ST27Q05 retained four categories. *HOMEPOS* was also one of three components in the construction of the *PISA index of economic, social and cultural status* (or ESCS; see the section on ESCS index construction later in this chapter). Table 16.2 shows the wording of items and their allocation to the four indices.

Table 16.2 **Household possessions and home background indices**

		Item is used to measure index			
Item		**WEALTH**	**CULTPOS**	**HEDRES**	**HOMEPOS**
ST26	**In your home, do you have:**				
ST26Q01	A desk to study at			X	X
ST26Q02	A room of your own	X			X
ST26Q03	A quiet place to study			X	X
ST26Q04	A computer you can use for school work			X	X
ST26Q05	Educational software			X	X
ST26Q06	A link to the Internet	X			X
ST26Q07	Classical literature		X		X
ST26Q08	Books of poetry		X		X
ST26Q09	Works of art		X		X
ST26Q10	Books to help with your school work			X	X
ST26Q11	Technical reference books			X	X
ST26Q12	A dictionary			X	X
ST26Q13	A dishwasher				X
ST26Q14	A <DVD> player	X			X
ST26Q15	<Country-specific wealth item 1>	X			X
ST26Q16	<Country-specific wealth item 2>	X			X
ST26Q17	<Country-specific wealth item 3>	X			X
ST27	**How many of these are there at your home?**				
ST27Q01	Cellular phones	X			X
ST27Q02	Televisions	X			X
ST27Q03	Computers	X			X
ST27Q04	Cars	X			X
ST27Q05	Rooms with a bath or shower	X			X
ST28	**How many books are there in your home?**				**X**

Table 16.3 provides information on the reliabilities in OECD countries for all four scales while Table 16.4 shows the corresponding reliabilities in partner countries and economies. Scale reliabilities were generally higher in partner countries and economies than in OECD countries. This may have been due to the higher degree of accessibility of household items for larger proportions of the population in economically more developed countries which saw these differentiate less between more and less affluent households in these countries. Table 16.5a provides international item parameters for home possession indices. Please note that items *ST26Q13- ST26Q17* were scaled nationally because these items had different meaning for different countries. National item parameters for these items are provided in Table 16.5b and were not included in the reliability scales to keep comparability of Cronbach's alpha across countries for each scale.

Table 16.3 **Scale reliabilities for home possession indices in OECD countries**

	HOMEPOS	CULTPOS	HEDRES	WEALTH
Australia	0.74	0.74	0.65	0.64
Austria	0.68	0.68	0.50	0.61
Belgium	0.70	0.70	0.53	0.62
Canada	0.71	0.71	0.61	0.63
Chile	0.84	0.84	0.57	0.80
Czech Republic	0.72	0.72	0.46	0.68
Denmark	0.71	0.71	0.52	0.59
Estonia	0.72	0.72	0.46	0.68
Finland	0.68	0.68	0.51	0.57
France	0.68	0.68	0.43	0.62
Germany	0.67	0.67	0.52	0.58
Greece	0.75	0.75	0.52	0.68
Hungary	0.76	0.76	0.49	0.69
Iceland	0.67	0.67	0.57	0.59
Ireland	0.70	0.70	0.56	0.59
Israel	0.76	0.76	0.56	0.75
Italy	0.70	0.70	0.44	0.61
Japan	0.67	0.67	0.48	0.55
Korea	0.77	0.77	0.55	0.64
Luxembourg	0.74	0.74	0.56	0.66
Mexico	0.87	0.87	0.60	0.85
Netherlands	0.64	0.64	0.52	0.53
New Zealand	0.74	0.74	0.66	0.62
Norway	0.71	0.71	0.59	0.62
Poland	0.75	0.75	0.48	0.71
Portugal	0.76	0.76	0.50	0.65
Slovak Republic	0.77	0.77	0.68	0.69
Slovenia	0.70	0.70	0.45	0.61
Spain	0.69	0.69	0.47	0.59
Sweden	0.71	0.71	0.55	0.61
Switzerland	0.64	0.64	0.49	0.55
Turkey	0.86	0.86	0.66	0.81
United Kingdom	0.72	0.72	0.62	0.61
United States	0.80	0.80	0.66	0.69
OECD median	**0.72**	**0.72**	**0.53**	**0.62**

Table 16.4 Scale reliabilities for home possession indices in partner countries and economies

	HOMEPOS	CULTPOS	HEDRES	WEALTH
Albania	0.81	0.81	0.59	0.76
Argentina	0.77	0.77	0.52	0.71
Brazil	0.80	0.80	0.54	0.75
Bulgaria	0.79	0.79	0.60	0.73
Colombia	0.82	0.82	0.58	0.76
Costa Rica	0.84	0.84	0.59	0.80
Croatia	0.71	0.71	0.44	0.63
Cyprus[1, 2]	0.73	0.73	0.56	0.63
Hong Kong-China	0.76	0.76	0.53	0.66
Indonesia	0.84	0.84	0.59	0.81
Jordan	0.84	0.84	0.70	0.79
Kazakhstan	0.82	0.82	0.64	0.77
Latvia	0.74	0.74	0.44	0.68
Liechtenstein	0.71	0.71	0.51	0.59
Lithuania	0.76	0.76	0.51	0.69
Macao-China	0.74	0.74	0.56	0.68
Malaysia	0.77	0.77	0.54	0.71
Montenegro	0.80	0.80	0.59	0.74
Peru	0.85	0.85	0.63	0.82
Qatar	0.78	0.78	0.70	0.77
Romania	0.82	0.82	0.61	0.75
Russian Federation	0.75	0.75	0.54	0.68
Serbia	0.77	0.77	0.47	0.71
Shanghai-China	0.81	0.81	0.52	0.75
Singapore	0.75	0.75	0.57	0.63
Chinese Taipei	0.75	0.75	0.60	0.63
Thailand	0.85	0.85	0.67	0.81
Tunisia	0.86	0.86	0.69	0.82
United Arab Emirates	0.78	0.78	0.65	0.77
Uruguay	0.79	0.79	0.57	0.72
Viet Nam	0.82	0.82	0.54	0.81
Median	**0.79**	**0.79**	**0.57**	**0.74**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Table 16.5a International item parameters for home possession indices

		Parameter estimates							
		WEALTH			CULTPOS	HEDRES	HOMEPOS		
Item		Delta	tau_1	tau_2	Delta	Delta	Delta	tau_1	tau_2
ST26	**In your home, do you have:**								
ST26Q01	A desk to study at					-0.92754	-1.54387		
ST26Q02	A room of your own	-1.27697					-0.79605		
ST26Q03	A quiet place to study					-0.48891	-1.15166		
ST26Q04	A computer you can use for school work					-0.20727	-0.81434		
ST26Q05	Educational software					2.09794	1.06883		
ST26Q06	A link to the Internet	-0.45176					-0.01091		
ST26Q07	Classical literature				0.08751		1.00802		
ST26Q08	Books of poetry				0.16730		1.07070		
ST26Q09	Works of art				0.25481		0.73601		
ST26Q10	Books to help with your school work					-0.29046	0.97589		
ST26Q11	Technical reference books					1.65251	0.80082		
ST26Q12	A dictionary					-1.83627	2.37387		
ST27	**How many of these are there at your home?**								
ST27Q01	Cellular phones	-0.44841	0.58740				0.08010	0.69876	
ST27Q02	Televisions	0.16215	0.50786				0.61400	0.39179	
ST27Q03	Computers	0.54215	-1.43777	0.57453			0.95064	1.24456	0.56661
ST27Q04	Cars	0.82013	-1.57244	0.02151			1.20761	-1.36576	0.01463
ST27Q05	Rooms with a bath or shower	0.65271	-2.85942	0.69222			1.03345	-2.59720	0.67423
ST28	**How many books are there in your home?**						**0.92900**	**-0.08203**	

Table 16.5b National item parameters for items ST26Q13-ST26Q17

		HOMEPOS					WEALTH				
		ST26Q13	ST26Q14	ST26Q15	ST26Q16	ST26Q17	ST26Q13	ST26Q14	ST26Q15	ST26Q16	ST26Q17
OECD	Australia	0.92001	-2.99451	1.52384	1.64010	1.72177	0.81937	-3.09018	1.49028	1.60871	1.69333
	Austria	-0.82504	-2.14069	0.44607	0.39620	0.72274	-1.20657	-2.44629	0.15728	0.10561	0.44421
	Belgium	-0.03350	-2.90529	0.90322	2.02001	1.94527	-0.30989	-3.18368	0.70531	1.83671	1.75994
	Canada	0.07628	-3.15817	-0.96986	1.12842	0.71438	-0.10286	-3.36528	-1.14859	1.03136	0.59748
	Chile	0.31967	-1.61146	-0.37854	0.09626	0.25152	-0.19587	-2.11183	-0.85569	-0.37000	-0.21043
	Czech Republic	1.68771	-1.90989	0.69060	1.01578	0.19373	0.97915	-2.60326	-0.00026	0.33711	-0.45745
	Denmark	-0.56525	-3.37462	1.34240	-1.40201	1.20628	-0.56408	-3.43123	1.28240	-1.45876	1.14635
	Estonia	2.07437	-0.49580	1.41945	-0.61023	0.14241	1.44800	-1.19011	0.77418	-1.30682	-0.53814
	Finland	-0.91111	-2.33719	-0.18326	0.72674	1.38458	-1.20925	-2.62107	-0.45258	0.46984	1.13760
	France	-0.06990	-2.97166	0.69948	0.82488	0.86548	-0.47749	-3.34517	0.41808	0.54822	0.59096
	Germany	-0.53598	-2.30400	0.30271	0.82117	1.30732	-0.93518	-2.67240	-0.01881	0.49996	0.99776
	Greece	0.32384	-2.19891	1.76390	0.59751	1.73481	-0.12290	-2.63569	1.44201	0.25407	1.41192
	Hungary	0.97335	-2.48418	-0.11226	0.77080	0.65854	0.21316	-3.16436	-0.82977	0.06388	-0.04951
	Iceland	-0.40061	-2.58105	2.86889	2.78256	1.45703	-0.90331	-3.04338	2.43042	2.34328	1.00722
	Ireland	-0.01310	-3.75455	-0.30069	0.89664	0.57593	-0.14915	-3.87541	-0.32222	0.89358	0.50171
	Israel	0.80221	-0.97503	2.60077	1.87352	1.87048	0.32855	-1.47998	2.27761	1.50927	1.51400
	Italy	0.99334	-2.30967	1.35900	1.70954	1.22048	0.58811	-2.70862	0.99389	1.34562	0.85501
	Japan	1.52598	-2.07939	-0.84910	-0.04520	0.59179	1.35315	-2.26600	-1.03420	-0.22853	0.41053
	Korea	2.06627	0.68528	0.82654	-0.01571	0.83472	1.32317	-0.01721	0.12305	-0.70282	0.12900
	Luxembourg	-1.11390	-2.46776	0.62207	0.92144	0.24069	-1.35117	-2.68419	0.47026	0.77427	0.08314
	Mexico	2.28317	-1.83565	0.30166	-0.53624	-0.46241	1.90490	-2.57429	-0.21010	-1.13442	-1.05248
	Netherlands	-0.22987	-3.00590	1.01001	2.74732	0.27238	-0.34611	-3.13882	0.90282	2.65832	0.15596
	New Zealand	0.48069	-2.57486	0.40991	0.59301	1.19121	0.33606	-2.71838	0.33609	0.52125	1.10131
	Norway	-1.24781	-2.66837	2.29501	2.51362	1.95438	-1.50608	-2.90709	2.08483	2.30569	1.71219
	Poland	2.20325	-1.28919	0.21525	0.16732	1.16861	1.41126	-2.10977	-0.62819	-0.67797	0.44005
	Portugal	0.55289	-1.37820	-0.28270	1.27435	2.44553	0.25755	-1.64679	-0.52775	1.07539	2.28063
	Slovak Republic	2.26815	-1.69047	1.29991	0.19004	0.30678	1.53405	-2.41324	0.60359	-0.52448	-0.40585
	Slovenia	0.06242	-1.32179	0.06507	0.93390	2.19644	-0.29738	-1.69151	-0.29459	0.58351	1.86380
	Spain	0.66871	-2.44119	-0.38198	0.89927	1.62306	0.23296	-2.86859	-0.78215	0.52507	1.26599
	Sweden	-0.25694	-3.19230	2.44263	1.32602	1.86250	-0.34381	-3.29017	2.36191	1.24096	1.77952
	Switzerland	-0.58744	-2.20113	0.41359	1.00812	0.56765	-0.88262	-2.49863	0.19503	0.76004	0.30900
	Turkey	0.03688	-0.45954	1.09665	0.55897	0.94418	-0.65294	-1.19760	0.40232	-0.13371	0.25030
	United Kingdom	0.82683	-3.20227	-0.18204	0.01550	1.72790	0.67016	-3.35310	-0.30782	-0.10822	1.62409
	United States	0.11783	-2.87797	1.80754	0.00510	0.74034	0.17057	-2.85219	1.90622	0.06490	0.81581
Partners	Albania	1.21681	-1.20501	-0.28034	-0.31013	-0.10923	0.57085	-1.86225	-0.90941	-0.94073	-0.73330
	Argentina	1.23191	-1.30960	0.34978	0.01160	-2.24498	0.64023	-1.92822	-0.21679	-0.56633	-2.88646
	Brazil	1.23447	-1.82940	-0.20238	0.25991	1.46311	0.75787	-2.42878	-0.71385	-0.22469	1.04050
	Bulgaria	1.82280	-0.84258	1.05605	-0.18339	1.34047	1.16793	-1.42923	0.49416	-0.76491	0.78444
	Colombia	-2.13670	-1.24874	-0.34186	-1.31106	-1.28262	-2.97053	-2.05642	-1.11871	-2.12079	-2.09167
	Costa Rica	1.10039	-1.73227	-0.12289	0.31990	0.91284	0.64121	-2.30074	-0.63377	-0.17240	0.44568
	Croatia	0.44244	-2.13399	0.81493	1.66983	0.97007	0.07634	-2.51804	0.45316	1.31809	0.61007
	Cyprus[1, 2]	0.21793	-1.33632	2.22184	-0.99834	2.86380	0.22663	-1.35277	2.28447	-1.01027	2.94143
	Hong Kong-China	3.18175	-2.00308	0.76515	0.97588	0.84677	2.56917	-2.56486	0.18740	0.39278	0.24136
	Indonesia	0.72581	-1.86522	0.62727	-1.62599	1.37452	-0.21394	-2.99129	-0.30063	-2.73399	0.50191
	Jordan	1.82083	-0.42988	1.13782	-0.34879	0.71787	1.43282	-0.88118	0.72637	-0.79844	0.29311
	Kazakhstan	2.52300	-2.05933	0.50627	1.12367	0.19343	1.60888	-3.07687	-0.45765	0.17641	-0.77869
	Latvia	2.78833	-0.74066	0.24415	0.87602	0.41383	1.83584	-1.67520	-0.65336	0.00646	-0.47617
	Liechtenstein	-1.26426	-2.48280	0.35808	0.83648	0.00816	-1.43671	-2.65339	0.25895	0.70073	-0.14055
	Lithuania	2.42182	-0.34677	0.47916	1.15563	0.90951	1.87083	-0.90557	-0.07859	0.59915	0.35261
	Macao-China	2.91259	-1.55873	0.28086	0.07277	-0.18504	2.59497	-1.95789	-0.08373	-0.29531	-0.55989
	Malaysia	2.42760	-1.03209	-4.07799	-3.02713	0.89650	2.04273	-1.64914	-4.79127	-3.72328	0.40900
	Montenegro	0.46432	-1.97117	-0.21968	1.35462	0.15915	-0.07288	-2.51741	-0.76032	0.82539	-0.37972
	Peru	3.25512	-1.70496	-0.76344	-0.40852	1.01574	2.54814	-2.86656	-1.83066	-1.43796	0.13374
	Qatar	1.46986	-0.62698	0.11477	-0.09098	1.12366	1.73788	-0.60044	0.21134	-0.01645	1.34451
	Romania	1.98883	-0.08632	0.86913	1.60458	1.83613	1.12649	-0.95401	0.03311	0.79409	1.03342
	Russian Federation	3.08632	-1.48144	0.20223	1.72097	1.63714	1.98126	-2.61474	-0.86323	0.71981	0.63337
	Serbia	1.53910	-0.99978	1.01031	2.16185	0.71458	0.97066	-1.59069	0.43540	1.60118	0.13637
	Shanghai-China	2.93689	-1.44542	0.65878	-0.18624	0.12192	2.25693	-2.18398	-0.05694	-0.91341	-0.60152
	Singapore	1.51213	-1.84961	-0.46405	-0.56857	2.35257	1.18476	-2.17985	-0.79381	-0.89849	2.02619
	Chinese Taipei	1.52702	-1.38425	1.34433	1.61435	0.62799	1.10773	-1.77959	0.92604	1.19479	0.21518
	Thailand	1.01137	-2.31523	1.34014	-1.09118	0.98234	0.31123	-3.19055	0.69982	-1.89802	0.31797
	Tunisia	1.47195	-0.77529	-0.35494	-0.70572	-1.40743	0.95317	-1.43989	-0.88738	-1.23852	-1.94124
	United Arab Emirates	1.46428	-1.09524	0.39164	0.32730	1.75656	1.63847	-1.26864	0.39432	0.31829	1.98312
	Uruguay	1.26025	-1.31281	-0.74818	-0.54775	0.19524	0.68181	-1.89856	-1.34948	-1.14790	-0.39735
	Viet Nam	0.48032	-1.21234	1.69718	-3.15495	2.60703	-0.34963	-2.18230	0.96541	-4.24295	1.93235

1. Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Attitudes towards mathematics

With mathematics being the major domain in 2012, attitudes towards this subject received considerable attention in the PISA 2012 Student Questionnaire. In total, ten indices were constructed using 67 items. Table 16.6 summarises all indices related to attitudes towards mathematics and their relationship to the similar indices in the PISA 2003 survey, the previous cycle in which mathematics was a major domain.

Table 16.6 **Attitudes towards mathematics indices**

Index	Index label	Relationship to other PISA surveys
INTMAT	Mathematics Interest	Used in 2003
INSTMOT	Instrumental Motivation for Mathematics	Used in 2003
SUBNORM	Subjective Norms in Mathematics	New
MATHEFF	Mathematics Self-Efficacy	Used in 2003
ANXMAT	Mathematics Anxiety	Used in 2003
SCMAT	Mathematics Self-Concept	Used in 2003
FAILMAT	Attributions to Failure in Mathematics	New
MATWKETH	Mathematics Work Ethic	New
MATINTFC	Mathematics Intentions	New
MATBEH	Mathematics Behaviour	New

Table 16.7 shows a high degree of internal consistency across OECD countries for all indices except Subjective Norms in Mathematics (*SUBNORM*) and Attributions to Failure in Mathematics (*FAILMAT*) which exhibit moderate to high reliability.

Table 16.7 **Scale reliabilities for attitudes towards mathematics indices in OECD countries**

	INTMAT	INSTMOT	SUBNORM	MATHEFF	ANXMAT	SCMAT	FAILMAT	MATWKETH	MATINTFC	MATBEH
Australia	0.91	0.90	0.69	0.88	0.85	0.89	0.71	0.91	0.76	0.77
Austria	0.88	0.86	0.70	0.82	0.86	0.90	0.62	0.87	0.76	0.70
Belgium	0.88	0.89	0.64	0.84	0.83	0.88	0.60	0.87	0.76	0.71
Canada	0.90	0.90	0.68	0.86	0.87	0.91	0.70	0.88	0.74	0.75
Chile	0.89	0.89	0.66	0.83	0.70	0.90	0.65	0.88	0.79	0.72
Czech Republic	0.87	0.88	0.63	0.81	0.85	0.91	0.56	0.84	0.77	0.64
Denmark	0.90	0.86	0.60	0.84	0.86	0.91	0.64	0.90	0.69	0.71
Estonia	0.89	0.87	0.66	0.82	0.86	0.89	0.61	0.83	0.70	0.67
Finland	0.90	0.89	0.71	0.85	0.82	0.92	0.68	0.88	0.83	0.72
France	0.87	0.88	0.65	0.82	0.77	0.90	0.57	0.89	0.71	0.71
Germany	0.89	0.84	0.63	0.81	0.87	0.91	0.64	0.87	0.77	0.70
Greece	0.89	0.90	0.67	0.83	0.81	0.87	0.56	0.89	0.73	0.76
Hungary	0.89	0.87	0.71	0.86	0.83	0.82	0.64	0.86	0.85	0.71
Iceland	0.91	0.91	0.70	0.88	0.88	0.91	0.73	0.91	0.69	0.78
Ireland	0.91	0.87	0.66	0.84	0.84	0.89	0.60	0.89	0.73	0.67
Israel	0.90	0.87	0.61	0.84	0.85	0.85	0.69	0.88	0.65*	0.78
Italy	0.87	0.89	0.63	0.81	0.78	0.88	0.59	0.87	0.65*	0.69
Japan	0.90	0.92	0.72	0.86	0.83	0.88	0.75	0.87	0.72	0.66
Korea	0.91	0.91	0.68	0.89	0.76	0.88	0.65	0.91	0.76	0.72
Luxembourg	0.89	0.91	0.70	0.86	0.84	0.90	0.70	0.89	0.73	0.81
Mexico	0.85	0.84	0.70	0.83	0.75	0.84	0.70	0.87	0.74	0.80
Netherlands	0.86	0.88	0.62	0.85	0.86	0.89	0.63	0.86	0.66	0.81
New Zealand	0.90	0.90	0.72	0.88	0.83	0.87	0.70	0.90	0.73	0.80
Norway	0.91	0.90	0.65	0.87	0.86	0.91	0.69	0.90	0.78	0.77
Poland	0.88	0.90	0.67	0.85	0.87	0.91	0.65	0.86	0.82	0.72
Portugal	0.87	0.91	0.68	0.88	0.78	0.89	0.62	0.91	0.74	0.74
Slovak Republic	0.88	0.87	0.70	0.83	0.84	0.83	0.63	0.88	0.79	0.75
Slovenia	0.89	0.88	0.68	0.85	0.83	0.86	0.70	0.87	0.80	0.79
Spain	0.86	0.90	0.62	0.82	0.79	0.89	0.58	0.88	0.76	0.71
Sweden	0.92	0.89	0.67	0.86	0.84	0.89	0.65	0.89	0.71	0.76
Switzerland	0.87	0.87	0.65	0.83	0.84	0.90	0.62	0.87	0.77	0.70
Turkey	0.89	0.87	0.71	0.82	0.82	0.85	0.66	0.91	0.77	0.80
United Kingdom	0.90	0.86	0.66	0.86	0.84	0.88	0.64	0.90	0.70	0.72
United States	0.91	0.91	0.71	0.85	0.88	0.90	0.73	0.88	0.76	0.80
OECD median	**0.89**	**0.89**	**0.67**	**0.85**	**0.84**	**0.89**	**0.64**	**0.88**	**0.75**	**0.72**

* Item ST48Q04 was deleted by the country.

Table 16.8 also shows high degree of internal consistency across partner countries for all indices except Attributions to Failure in Mathematics (*FAILMAT*) which exhibits moderate to high reliability.

Table 16.8 Scale reliabilities for attitudes towards mathematics indices in partner countries and economies

	INTMAT	INSTMOT	SUBNORM	MATHEFF	ANXMAT	SCMAT	FAILMAT	MATWKETH	MATINTFC	MATBEH
Albania	0.85	0.84	0.72	0.81	0.79	0.81	0.60	0.90	0.78	0.80
Argentina	0.88	0.85	0.67	0.81	0.75	0.83	0.64	0.89	0.77	0.81
Brazil	0.84	0.84	0.69	0.82	0.71	0.82	0.63	0.87	0.77	0.82
Bulgaria	0.89	0.88	0.76	0.84	0.83	0.83	0.66	0.90	0.79	0.86
Colombia	0.84	0.86	0.68	0.82	0.72	0.83	0.66	0.88	0.80	0.78
Costa Rica	0.88	0.87	0.69	0.79	0.77	0.85	0.65	0.86	0.77	0.74
Croatia	0.90	0.88	0.68	0.83	0.84	0.88	0.63	0.88	0.75	0.72
Cyprus[1, 2]	0.91	0.90	0.70	0.86	0.82	0.86	0.72	0.91	0.72	0.82
Hong Kong-China	0.91	0.90	0.73	0.89	0.85	0.89	0.71	0.90	0.76	0.77
Indonesia	0.83	0.79	0.76	0.82	0.74	0.73	0.64	0.88	0.78	0.84
Jordan	0.84	0.87	0.73	0.83	0.51	0.80	0.74	0.90	0.61	0.84
Kazakhstan	0.89	0.87	0.81	0.84	0.82	0.82	0.77	0.90	0.80	0.83
Latvia	0.85	0.86	0.66	0.79	0.80	0.88	0.62	0.83	0.73	0.71
Liechtenstein	0.90	0.87	0.60	0.82	0.82	0.90	0.52	0.86	0.80	0.71
Lithuania	0.88	0.90	0.68	0.82	0.81	0.87	0.68	0.87	0.73*	0.74
Macao-China	0.90	0.90	0.72	0.86	0.86	0.90	0.70	0.86	0.72	0.75
Malaysia	0.87	0.85	0.74	0.80	0.74	0.81	0.70	0.88	0.72	0.80
Montenegro	0.91	0.91	0.78	0.85	0.79	0.86	0.66	0.91	0.85	0.83
Peru	0.87	0.85	0.68	0.80	0.69	0.81	0.65	0.90	0.75	0.79
Qatar	0.88	0.89	0.80	0.88	0.83	0.74	0.80	0.91	0.69	0.87
Romania	0.77	0.84	0.66	0.79	0.72	0.70	0.64	0.84	0.80	0.89
Russian Federation	0.86	0.90	0.73	0.83	0.80	0.83	0.63	0.89	0.84	0.74
Serbia	0.89	0.88	0.70	0.84	0.81	0.87	0.66	0.88	0.83	0.73
Shanghai-China	0.91	0.88	0.72	0.91	0.86	0.87	0.71	0.91	0.80	0.76
Singapore	0.90	0.87	0.67	0.86	0.83	0.88	0.71	0.88	0.73	0.73
Chinese Taipei	0.91	0.91	0.74	0.91	0.82	0.91	0.76	0.92	0.78	0.76
Thailand	0.86	0.84	0.75	0.82	0.63	0.80	0.76	0.86	0.71	0.79
Tunisia	0.87	0.86	0.68	0.79	0.68	0.85	0.62	0.88	0.79	0.80
United Arab Emirates	0.87	0.87	0.71	0.82	0.79	0.82	0.71	0.89	0.65	0.83
Uruguay	0.88	0.89	0.71	0.82	0.79	0.89	0.66	0.89	0.81	0.77
Viet Nam	0.82	0.81	0.67	0.78	0.71	0.77	0.45	0.85	0.85	0.70
Median	**0.88**	**0.87**	**0.71**	**0.82**	**0.79**	**0.83**	**0.66**	**0.88**	**0.77**	**0.79**

* Item ST48Q04 was deleted by the country.

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Four items measuring mathematics interest (*INTMAT*) were used in the Main Survey of PISA 2012 as well as for PISA 2003. Table 16.9 shows the item wording and the international item parameters for this scale. The response categories were "Strongly agree", "Agree", "Disagree" and "Strongly disagree". All items were reversed so that the higher difficulty corresponds to a higher level of interest. For this index, item difficulties vary from a comparatively easy one "I am interested in the things I learn in mathematics" ($\delta = -0.87$) to a more difficult "I enjoy reading about mathematics" ($\delta = 0.52$). This indicates that even those students who are interested in the things they learn in mathematics do not necessarily read about mathematics.

Table 16.9 Item parameters for mathematics interest (*INTMAT*)

Item	Thinking about your views on mathematics: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST29Q01	a) I enjoy reading about mathematics	0.51591	-3.61604	0.04022
ST29Q03	c) I look forward to my mathematics lessons	0.28224	-3.63880	0.05071
ST29Q04	d) I do mathematics because I enjoy it	0.07060	-3.48903	0.14803
ST29Q06	f) I am interested in the things I learn in mathematics	-0.86875	-3.66945	-0.16382

Four items measuring instrumental motivation for mathematics (*INSTMOT*) were used in the Main Survey of PISA 2012 as well as for PISA 2003. Table 16.10 shows the item wording and the international item parameters for this scale. The response categories vary from "Strongly Agree", "Agree", "Disagree", to "Strongly disagree". All items were reversed, so the higher difficulty corresponds to the higher level of motivation. For this index, item difficulties do not vary considerably.

Table 16.10 Item parameters for instrumental motivation for mathematics (*INSTMOT*)

Item	Thinking about your views on mathematics: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST29Q02	b) Making an effort in mathematics is worth it because it will help me in the work that I want to do later on	-0.21392	-2.92583	-0.54753
ST29Q05	e) Learning mathematics is worthwhile for me because it will improve my career <prospects, chances>	-0.26314	-2.80478	-0.65997
ST29Q07	g) Mathematics is an important subject for me because I need it for what I want to study later on	0.29837	-2.76227	-0.34159
ST29Q08	h) I will learn many things in mathematics that will help me get a job	0.17869	-2.95336	-0.51783

A new scale was created in PISA 2012 consisting of six items measuring subjective norms in mathematics (*SUBNORM*) in the Main Survey. Table 16.11 shows the item wording and the international item parameters for this scale. The response categories range from "Strongly agree", "Agree", "Disagree" to "Strongly disagree". All items were reversed, so the higher difficulty corresponds to the higher level of agreement. Values for delta indicated that students found it harder to agree with the statement "My friends enjoy taking mathematics tests" ($\delta = 1.36$) whereas it was relatively easy to agree with the statement "My parents believe it's important for me to study mathematics" ($\delta = -1.08$).

Table 16.11 Item parameters for subjective norms in mathematics (*SUBNORM*)

Item	Thinking about how people important to you view mathematics: how strongly do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST35Q01	a) Most of my friends do well in mathematics	0.08571	-2.48453	-0.14556
ST35Q02	b) Most of my friends work hard at mathematics	0.27374	-2.47999	0.06737
ST35Q03	c) My friends enjoy taking mathematics tests	1.36375	-2.00831	0.53358
ST35Q04	d) My parents believe it's important for me to study mathematics	-1.08125	-1.20914	-0.77788
ST35Q05	e) My parents believe that mathematics is important for my career	-0.77458	-1.58049	-0.24409
ST35Q06	f) My parents like mathematics	0.13263	-1.97702	-0.13822

Eight items measuring mathematics self-efficacy (*MATHEFF*) were used in the Main Survey of PISA 2012 as well as for PISA 2003. Table 16.12 shows the item wording and the international item parameters for this scale. The response categories were "Very confident", "Confident", "Not very confident" and "Not at all confident". All items were reversed, so the higher difficulty corresponds to the higher level of confidence. For this index, item difficulties ranged from a comparatively easy one "Solving an equation like 3x+5= 17" ($\delta = -0.62$) to more difficult ones, such as "Finding the actual distance between two places on a map with a 1:10 000 scale" ($\delta = 0.56$) and "Calculating the petrol consumption rate of a car" ($\delta = 0.65$). This indicates that students felt more confident in solving linear equations than they felt applying rates and proportions to real life situations.

Table 16.12 Item parameters for mathematics self-efficacy (*MATHEFF*)

Item	How confident do you feel about having to do the following mathematics tasks?	Parameter estimates		
		Delta	tau_1	tau_2
ST37Q01	a) Using a <train timetable> to work out how long it would take to get from one place to another	-0.16537	-1.85539	-0.10055
ST37Q02	b) Calculating how much cheaper a TV would be after a 30% discount	-0.31699	-1.69833	-0.05907
ST37Q03	c) Calculating how many square metres of tiles you need to cover a floor	0.13366	-1.83729	0.14817
ST37Q04	d) Understanding graphs presented in newspapers	-0.15505	-1.78957	-0.12065
ST37Q05	e) Solving an equation like 3x+5= 17	-0.62211	-1.09181	-0.05693
ST37Q06	f) Finding the actual distance between two places on a map with a 1:10 000 scale	0.55600	-1.89647	0.29021
ST37Q07	g) Solving an equation like 2(x+3) = (x + 3) (x - 3)	-0.07740	-1.38321	0.06433
ST37Q08	h) Calculating the petrol consumption rate of a car	0.64726	-1.96788	0.13774

The five items measuring mathematics anxiety (*ANXMAT*) that were used in the Main Survey of PISA 2012 had also been used in PISA 2003. Table 16.13 shows the item wording and the international item parameters for this scale. The response categories were "strongly agree", "agree", "disagree" and "strongly disagree". All items were reversed, so the higher difficulty corresponds to the higher level of anxiety.

Table 16.13 **Item parameters for mathematics anxiety (*ANXMAT*)**

Item	Thinking about studying mathematics: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST42Q01	a) I often worry that it will be difficult for me in mathematics classes	-0.66180	-2.01172	-0.09668
ST42Q03	c) I get very tense when I have to do mathematics homework	0.38199	-2.13733	0.37662
ST42Q05	e) I get very nervous doing mathematics problems	0.41758	-2.30251	0.35980
ST42Q08	h) I feel helpless when doing a mathematics problem	0.55522	-2.20650	0.41807
ST42Q10	j) I worry that I will get poor <grades> in mathematics	-0.69299	-1.42498	-0.23096

Five items measuring mathematics self-concept (*SCMAT*) were used in the Main Survey of PISA 2012 as well as in PISA 2003. Table 16.14 shows the item wording and the international item parameters for this scale. The response categories range from “Strongly agree” to “Strongly disagree”. All items except ST42Q02 were reversed, so the higher difficulty corresponds to the higher self-concept.

Table 16.14 **Item parameters for mathematics self-concept (*SCMAT*)**

Item	Thinking about studying mathematics: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST42Q02	b) I am just not good at mathematics	-0.14882	-2.13451	-0.24910
ST42Q04	d) I get good <grades> in mathematics	-0.48764	-2.75557	-0.08299
ST42Q06	f) I learn mathematics quickly	-0.22529	-2.75319	0.01990
ST42Q07	g) I have always believed that mathematics is one of my best subjects	0.32326	-2.02405	0.24282
ST42Q09	i) In my mathematics class, I understand even the most difficult work	0.53849	-2.69926	0.12835

A new scale indicating attributions to failure in mathematics (*FAILMAT*) was developed in PISA 2012 based on six items. Table 16.15 shows the item wording and the international item parameters for this scale. The response categories were “Very likely”, “Likely”, “Slightly likely”, “Not at all likely”. All items were reversed, so the higher difficulty corresponds to the higher level of external attribution of failure such as bad luck, bad guesses or the teacher.

Table 16.15 **Item parameters for attributions to failure in mathematics (*FAILMAT*)**

Item	Suppose that you are a student in the following situation: *Each week, your mathematics teacher gives a short quiz. Recently you have done badly on these quizzes. Today you are trying to figure out why.* How likely are you to have these thoughts or feelings in this situation?	Parameter estimates		
		Delta	tau_1	tau_2
ST44Q01	a) I’m not very good at solving mathematics problems	-0.13633	-1.24959	-0.27963
ST44Q03	b) My teacher did not explain the concepts well this week	0.24787	-1.14744	-0.10283
ST44Q04	c) This week I made bad guesses on the quiz	0.28440	-0.99882	-0.29146
ST44Q05	d) Sometimes the course material is too hard	-0.55077	-1.04547	-0.29291
ST44Q07	e) The teacher did not get students interested in the material	0.04449	-0.96201	0.00242
ST44Q08	f) Sometimes I am just unlucky	0.11034	-0.76450	-0.22261

Nine items were used in the Main Survey of PISA 2012 to create a new scale labeled “Mathematics work ethic” (*MATWKETH*). Table 16.16 shows the item wording and the international item parameters for this scale. The response categories ranged from “Strongly agree” to “Strongly disagree”. All items were reversed, so the higher difficulty corresponds to the higher level of work ethic.

Table 16.16 **Item parameters for mathematics work ethic (*MATWKETH*)**

Item	Thinking about the mathematics you do for school: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST46Q01	a) I finish my homework in time for mathematics class	-0.05155	-2.11022	-0.25487
ST46Q02	b) I work hard on my mathematics homework	0.23278	-2.46941	-0.07761
ST46Q03	c) I am prepared for my mathematics exams	0.06373	-2.39979	-0.16623
ST46Q04	d) I study hard for mathematics quizzes	0.44044	-2.57064	0.07100
ST46Q05	e) I keep studying until I understand mathematics material	0.13279	-2.53385	-0.02092
ST46Q06	f) I pay attention in mathematics class	-0.44658	-2.29932	-0.48632
ST46Q07	g) I listen in mathematics class	-0.69306	-2.12994	-0.72488
ST46Q08	h) I avoid distractions when I am studying mathematics	0.11249	-2.59695	-0.03840
ST46Q09	i) I keep my mathematics work well organised	0.20896	-2.50181	-0.04529

The five items measuring mathematics intentions (*MATINTFC*) are all of the so-called “Forced Choice” format which was one of the new item types employed in PISA 2012 (see also Chapters 3 and 17). This item type forces students to choose between mathematics and some other subject like language or science with respect to additional courses at school and

beyond. All items were reversed. Table 16.17 shows the item wording and the international item parameters for this scale. The item deltas show that the items forcing students to choose between mathematics and the test language are easier than the items that force students to choose between mathematics and science.

Table 16.17 Item parameters for mathematics intentions (*MATINTFC*)

Item	For each pair of statements, please choose the item that best describes you	Parameter estimates
		Delta
ST48Q01	a) 1. I intend to take additional mathematics courses after school finishes	-0.50884
	2. I intend to take additional <test language> courses after school finishes	
ST48Q02	b) 1. I plan on majoring in a subject in <college> that requires mathematics skills	0.53523
	2. I plan on majoring in a subject in <college> that requires science skills	
ST48Q03	c) 1. I am willing to study harder in my mathematics classes than is required	-0.53939
	2. I am willing to study harder in my <test language> classes than is required	
ST48Q04	d) 1. I plan on <taking> as many mathematics classes as I can during my education	0.05019
	2. I plan on <taking> as many science classes as I can during my education	
ST48Q05	e) 1. I am planning on pursuing a career that involves a lot of mathematics	0.46281
	2. I am planning on pursuing a career that involves a lot of science	

Another new scale was created in PISA 2012 to indicate students mathematics behaviour regarding mathematics both at and outside school (*MATBEH*) based on eight items. Table 16.18 shows the item wording and the international item parameters for this scale. The response categories were “Always or almost always”, “Often”, “Sometimes” and “Never or rarely”. All items were reversed, so the higher difficulty corresponds to the higher frequency. For this index, most of the items were easy, indicating, for example, that students frequently talk to their friends about mathematics ($\delta = -0.44$) and help their friends with mathematics ($\delta = -0.61$). However, participation in mathematics clubs or mathematics competitions is comparatively rare, probably because it depends on whether such activities are offered by the school.

Table 16.18 Item parameters for mathematics behaviour (*MATBEH*)

Item	How often do you do the following things at school and outside of school?	Parameter estimates		
		Delta	tau_1	tau_2
ST49Q01	a) I talk about mathematics problems with my friends	-0.43912	-1.36979	0.28510
ST49Q02	b) I help my friends with mathematics	-0.61070	-1.62219	0.24023
ST49Q03	c) I do mathematics as an <extracurricular> activity	-0.06138	-0.86084	0.08992
ST49Q04	d) I take part in mathematics competitions	0.32585	-0.23329	-0.06082
ST49Q05	e) I do mathematics more than 2 hours a day outside of school	0.13873	-0.79764	0.25319
ST49Q06	f) I play chess	-0.00254	-0.53372	0.04487
ST49Q07	g) I program computers	-0.07083	-0.36567	-0.06427
ST49Q09	h) I participate in a mathematics club	0.71999	0.41950	-0.42667

Opportunity to learn (OTL)

In PISA 2012 opportunity to learn scales included indices relating to student-perceived experiences and familiarity with mathematical tasks (OTL content: *EXAPPLM*, *EXPUREM*, *FAMCON*, *FAMCONC*) as well as indices relating to student-perceived teaching practices (*TCHBEHTD*, *TCHBEHFA*, *TCHBEHSO*) and student-perceived teaching quality (*TEACHSUP*, *COGACT*, *MTSUP*, *CLSMAN*, *DISCLIMA*). Table 16.19 summarises all scaled indices related to opportunity to learn.

Table 16.19 Opportunity to learn indices

Index	Index label	Relationship to other PISA surveys
OTL - Content		
EXAPPLM	Experience with Applied Mathematics Tasks at School	New
EXPUREM	Experience with Pure Mathematics Tasks at School	New
FAMCON	Familiarity with Mathematical Concepts	New
FAMCONC	Familiarity with Mathematical Concepts (Signal Detection Adjusted)	New composite scale based on FAMCONC (see description above Table 16.24)
OTL - Teaching Practices		
TCHBEHTD	Teacher Behaviour: Teacher-directed Instruction	New
TCHBEHFA	Teacher Behaviour: Formative Assessment	New
TCHBEHSO	Teacher Behaviour: Student Orientation	New
OTL - Teaching Quality		
TEACHSUP	Math Teaching	Used in 2003
COGACT	Cognitive Activation	New
DISCLIMA	Disciplinary Climate	Similar index was used in 2000 and 2009, modified in 2012
MTSUP	Teacher Support	New
CLSMAN	Classroom Management	New

Table 16.20 shows a high degree of internal consistency across OECD countries for all indices except teacher behavior - student orientation index (*TCHBEHSO*) which exhibit moderate to high reliability.

[Part 1/2]
Table 16.20 Scale reliabilities and correlations for opportunity to learn indices in OECD countries

	Cronbach's Alpha							Correlation between
	EXAPPLM	EXPUREM	FAMCON	TCHBEHTD	TCHBEHFA	TCHBEHSO	TEACHSUP	TCHBEHSO and TCHBEHFA
Australia	0.78	0.92	0.90	0.76	0.79	0.69	0.90	0.49
Austria	0.69	0.94	0.84	0.69	0.73	0.68	0.84	0.43
Belgium	0.74	0.94	0.86	0.67	0.72	0.67	0.84	0.44
Canada	0.77	0.94	0.84	0.76	0.78	0.69	0.88	0.50
Chile	0.77	0.93	0.87	0.74	0.77	0.67	0.87	0.53
Czech Republic	0.73	0.89	0.84	0.67	0.67	0.51	0.83	0.44
Denmark	0.81	0.88	0.84	0.70	0.77	0.60	0.84	0.46
Estonia	0.71	0.91	0.82	0.68	0.73	0.63	0.83	0.49
Finland	0.76	0.90	0.85	0.74	0.78	0.64	0.87	0.49
France	0.71	0.92	0.82	0.71	0.71	0.63	0.84	0.38
Germany	0.72	0.90	0.84	0.68	0.73	0.67	0.84	0.47
Greece	0.78	0.92	0.87	0.73	0.75	0.79	0.82	0.51
Hungary	0.75	0.88	0.86	0.73	0.75	0.72	0.85	0.50
Iceland	0.82	0.95	0.86	0.70	0.78	0.67	0.88	0.55
Ireland	0.73	0.93	0.85	0.73	0.75	0.63	0.87	0.35
Israel	0.77	0.88	0.83	0.71	0.76	0.64	0.85	0.56
Italy	0.71	0.88	0.84	0.71	0.72	0.63	0.84	0.47
Japan	0.79	0.93	0.83	0.68	0.68	0.57	0.87	0.48
Korea	0.81	0.95	0.89	0.74	0.79	0.76	0.84	0.61
Luxembourg	0.75	0.93	0.86	0.74	0.72	0.76	0.87	0.54
Mexico	0.79	0.92	0.90	0.73	0.76	0.71	0.87	0.58
Netherlands	0.77	0.93	0.85	0.74	0.74	0.75	0.82	0.56
New Zealand	0.82	0.93	0.90	0.76	0.80	0.69	0.89	0.54
Norway	0.78	0.91	*	0.75	0.81	0.63	0.87	0.54
Poland	0.72	0.84	0.85	0.71	0.76	0.72	0.86	0.54
Portugal	0.82	0.95	0.85	0.78	0.81	0.79	0.90	0.59
Slovak Republic	0.75	0.84**	0.83	0.69	0.70	0.69	0.83	0.53
Slovenia	0.76	0.90	0.84	0.75	0.75	0.80	0.82	0.51
Spain	0.74	0.90	0.86	0.70	0.78	0.72	0.88	0.49
Sweden	0.81	0.90	0.91	0.75	0.79	0.67	0.89	0.61
Switzerland	0.71	0.92	0.81	0.68	0.73	0.67	0.82	0.52
Turkey	0.80	0.92	0.87	0.79	0.75	0.75	0.85	0.58
United Kingdom	0.77	0.93	0.88	0.71	0.79	0.62	0.87	0.47
United States	0.79	0.90	0.89	0.76	0.79	0.68	0.87	0.52
OECD median α	**0.77**	**0.92**	**0.85**	**0.73**	**0.76**	**0.68**	**0.85**	
OECD average correlation								**0.51**
Correlation S.D.								**0.06**

* Unit ST62 was deleted by the country.
** Item ST61Q09 was deleted during adjudication because of the printing error.

[Part 2/2]
Table 16.20 Scale reliabilities and correlations for opportunity to learn indices in OECD countries

	Cronbach's Alpha				Correlation between
	COGACT	MTSUP	CLSMAN	DISCLIMA	DISCLIMA and CLSMAN
Australia	0.88	0.85	0.76	0.91	0.64
Austria	0.81	0.67	0.71	0.89	0.66
Belgium	0.83	0.71	0.75	0.89	0.58
Canada	0.86	0.82	0.75	0.89	0.57
Chile	0.82	0.76	0.73	0.87	0.56
Czech Republic	0.81	0.77	0.75	0.90	0.60
Denmark	0.81	0.75	0.72	0.88	0.60
Estonia	0.81	0.75	0.75	0.90	0.59
Finland	0.85	0.74	0.72	0.90	0.62
France	0.80	0.74	0.79	0.89	0.66
Germany	0.79	0.68	0.71	0.89	0.63
Greece	0.80	0.77	0.71	0.82	0.48
Hungary	0.79	0.77	0.72	0.91	0.60
Iceland	0.89	0.83	0.75	0.89	0.55
Ireland	0.84	0.81	0.79	0.91	0.68
Israel	0.81	0.77	0.71	0.90	0.60
Italy	0.79	0.77	0.72	0.87	0.57
Japan	0.83	0.77	0.31	0.85	0.32
Korea	0.85	0.81	0.62	0.87	0.50
Luxembourg	0.85	0.77	0.72	0.88	0.60
Mexico	0.82	0.73	0.66	0.83	0.42
Netherlands	0.86	0.69	0.75	0.89	0.62
New Zealand	0.89	0.83	0.74	0.90	0.57
Norway	0.88	0.79	0.70	0.88	0.53
Poland	0.83	0.82	0.75	0.90	0.61
Portugal	0.90	0.81	0.74	0.90	0.62
Slovak Republic	0.80	0.71	0.65	0.86	0.48
Slovenia	0.82	0.76	0.65	0.90	0.50
Spain	0.83	0.79	0.70	0.89	0.52
Sweden	0.89	0.80	0.70	0.88	0.55
Switzerland	0.79	0.72	0.72	0.88	0.62
Turkey	0.83	0.81	0.57	0.85	0.38
United Kingdom	0.86	0.83	0.77	0.91	0.64
United States	0.87	0.84	0.75	0.89	0.47
OECD median α	**0.83**	**0.77**	**0.72**	**0.89**	
OECD average correlation					**0.56**
Correlation S.D.					**0.08**

* Unit ST62 was deleted by the country.
** Item ST61Q09 was deleted during adjudication because of the printing error.

The table also displays correlations between scales that were assumed to be related. Thus, teacher behaviour that was perceived to be more student oriented (*TCHBEHSO*) was expected to be positively related to teachers using formative assessments (*TCHBEHFA*) to guide further instruction. Similarly, good classroom management (*CLSMAN*) was assumed to be related to a better disciplinary climate (*DISCLIMA*). As can be seen, correlations between the indices *TCHBEHSO* and *TCHBEHFA* were positive (0.51) and quite consistent across countries as indicated by the small standard deviation (0.06). In addition, the average correlation (0.56) between *DISCLIMA* and *CLSMAN* was also reasonably consistent across countries as indicated by the standard deviation of 0.08. Given the coding of the variables underlying the *DISCLIMA* and

CLSMAN indices, the positive correlation indicated that where students reported fewer problems with the disciplinary climate, students were more likely to agree that there was good classroom management whereby teachers started lessons on time, kept the class orderly and got students to listen to them.

Table 16.21 shows that internal consistency for the opportunity to learn across partner countries and economies was slightly lower than for OECD countries with the classroom management index (CLSMAN) exhibiting the lowest but still acceptable median reliability (0.66). The correlations between *TCHBEHSO* and *TCHBEHFA* for partner countries and economies are similar to those in the OECD countries while a slightly lower median correlation (0.43) and larger standard deviation (0.11) is recorded for the correlation between *DISCLIMA* and *CLSMAN*, probably as a consequence of the lower reliability of *CLSMAN* in partner countries and economies.

[Part 1/2]

Table 16.21 Scale reliabilities and correlations for opportunity to learn indices in partner countries and economies

	Cronbach's Alpha							Correlation between
	EXAPPLM	**EXPUREM**	**FAMCON**	**TCHBEHTD**	**TCHBEHFA**	**TCHBEHSO**	**TEACHSUP**	**TCHBEHSO and TCHBEHFA**
Albania	0.73	0.87	0.91	0.63	0.70	0.68	0.70	0.56
Argentina	0.78	0.90	0.86	0.69	0.72	0.70	0.85	0.58
Brazil	0.76	0.87	0.90	0.71	0.76	0.67	0.82	0.54
Bulgaria	0.80	0.88	0.89	0.73	0.76	0.76	0.82	0.59
Colombia	0.76	0.91	0.88	0.70	0.74	0.59	0.82	0.54
Costa Rica	0.75	0.92	0.86	0.72	0.75	0.62	0.86	0.54
Croatia	0.73	0.88	0.83	0.74	0.70	0.70	0.81	0.49
Cyprus[1, 2]	0.83	0.93	0.86	0.77	0.78	0.76	0.87	0.56
Hong Kong-China	0.77	0.92	0.86	0.76	0.76	0.78	0.85	0.50
Indonesia	0.79	0.90	0.89	0.69	0.65	0.63	0.65	0.56
Jordan	0.80	0.84	0.95	0.81	0.78	0.80	0.88	0.70
Kazakhstan	0.75	0.85	0.89	0.69	0.65	0.67	0.76	0.51
Latvia	0.74	0.86	0.82	0.69	0.73	0.67	0.79	0.54
Liechtenstein	0.71	0.94	0.88	0.69	0.73	0.55	0.81	0.41
Lithuania	0.75	0.86	0.81	0.69	0.77	0.69	0.81	0.51
Macao-China	0.75	0.89	0.84	0.74	0.77	0.62	0.83	0.49
Malaysia	0.77	0.92	0.86	0.67	0.72	0.67	0.79	0.54
Montenegro	0.79	0.87	0.87	0.75	0.71	0.75	0.86	0.59
Peru	0.76	0.89	0.91	0.72	0.71	0.65	0.81	0.55
Qatar	0.84	0.89	0.91	0.79	0.79	0.79	0.86	0.70
Romania	0.76	0.84	0.91	0.68	0.69	0.77	0.77	0.56
Russian Federation	0.77	0.87	0.83	0.71	0.69	0.69	0.79	0.56
Serbia	0.78	0.89	0.86	0.75	0.75	0.74	0.86	0.60
Shanghai-China	0.81	0.92	0.81	0.70	0.69	0.75	0.81	0.56
Singapore	0.74	0.91	0.89	0.73	0.79	0.74	0.87	0.52
Chinese Taipei	0.81	0.97	0.88	0.78	0.74	0.69	0.86	0.49
Thailand	0.78	0.91	0.89	0.72	0.75	0.80	0.82	0.70
Tunisia	0.70	0.83	0.82	0.74	0.71	0.69	0.81	0.56
United Arab Emirates	0.78	0.89	0.89	0.75	0.77	0.73	0.85	0.62
Uruguay	0.79	0.91	0.87	0.68	0.72	0.68	0.86	0.53
Viet Nam	0.73	0.84	0.85	0.59	0.64	0.64	0.73	0.50
Median α	**0.77**	**0.89**	**0.87**	**0.72**	**0.73**	**0.69**	**0.82**	
Average correlation								**0.56**
Correlation S.D.								**0.06**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table 16.21 Scale reliabilities and correlations for opportunity to learn indices in partner countries and economies

	Cronbach's Alpha				Correlation between
	COGACT	MTSUP	CLSMAN	DISCLIMA	DISCLIMA and CLSMAN
Albania	0.74	0.75	0.54	0.85	0.35
Argentina	0.81	0.74	0.59	0.82	0.37
Brazil	0.82	0.78	0.64	0.84	0.42
Bulgaria	0.84	0.80	0.69	0.84	0.42
Colombia	0.80	0.73	0.68	0.81	0.41
Costa Rica	0.79	0.77	0.70	0.80	0.39
Croatia	0.81	0.82	0.74	0.90	0.55
Cyprus[1, 2]	0.84	0.83	0.68	0.85	0.44
Hong Kong-China	0.86	0.86	0.75	0.90	0.57
Indonesia	0.80	0.71	0.49	0.80	0.19
Jordan	0.86	0.82	0.54	0.84	0.31
Kazakhstan	0.80	0.79	0.62	0.85	0.42
Latvia	0.77	0.79	0.68	0.88	0.55
Liechtenstein	0.74	0.63	0.72	0.88	0.66
Lithuania	0.82	0.78	0.66	0.91	0.58
Macao-China	0.83	0.82	0.72	0.84	0.51
Malaysia	0.78	0.77	0.57	0.82	0.26
Montenegro	0.83	0.79	0.66	0.87	0.45
Peru	0.81	0.74	0.61	0.78	0.37
Qatar	0.88	0.82	0.48	0.89	0.35
Romania	0.77	0.70	0.60	0.84	0.44
Russian Federation	0.80	0.81	0.57	0.90	0.43
Serbia	0.82	0.77	0.72	0.89	0.49
Shanghai-China	0.84	0.82	0.71	0.88	0.54
Singapore	0.87	0.82	0.69	0.90	0.54
Chinese Taipei	0.85	0.88	0.73	0.91	0.56
Thailand	0.81	0.81	0.51	0.83	0.30
Tunisia	0.80	0.76	0.61	0.77	0.30
United Arab Emirates	0.83	0.82	0.64	0.86	0.38
Uruguay	0.80	0.76	0.66	0.88	0.42
Viet Nam	0.73	0.64	0.60	0.76	0.38
Median α	**0.81**	**0.79**	**0.66**	**0.85**	
Average correlation					**0.43**
Correlation S.D.					**0.11**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Six items measuring students' exposure to applied mathematics tasks at school (*EXAPPLM*) were used in PISA 2012. Table 16.22 shows the item wording and the international item parameters for this scale. Response categories were "Frequently", "Sometimes", "Rarely" and "Never". All items were reversed so the higher difficulty corresponds to the higher frequency of exposure to applied mathematics tasks.

Table 16.22 Item parameters for experience with applied mathematics tasks at school (*EXAPPLM*)

Item	How often have you encountered the following types of mathematics tasks during your time at school?	Parameter estimates		
		Delta	tau_1	tau_2
ST61Q01	a) Working out from a <train timetable> how long it would take to get from one place to another	-0.07187	-1.07751	-0.37378
ST61Q02	b) Calculating how much more expensive a computer would be after adding tax	0.02359	-0.97187	-0.37078
ST61Q03	c) Calculating how many square metres of tiles you need to cover a floor	-0.21670	-0.84392	-0.35893
ST61Q04	d) Understanding scientific tables presented in an article	0.10966	-1.10943	-0.22095
ST61Q06	f) Finding the actual distance between two places on a map with a 1:10 000 scale	-0.11147	-1.34065	-0.01261
ST61Q08	h) Calculating the power consumption of an electronic appliance per week	0.26679	-1.22494	-0.11481

Three items measuring experience with applied mathematics tasks at school (*EXPUREM*) were used in the Main Survey of PISA 2012. Table 16.23 shows the item wording and the international item parameters for this scale. Response categories were: "Frequently", "Sometimes", "Rarely", "Never". All items were reversed so the higher difficulty corresponds to the higher frequency of exposure to pure mathematics tasks at school. However all three items have similar difficulty around 0 indicating that all of them on average are similarly frequently experienced at school.

Table 16.23 Item parameters for experience with pure mathematics tasks at school (*EXPUREM*)

Item	How often have you encountered the following types of mathematics tasks during your time at school?	Parameter estimates		
		Delta	tau_1	tau_2
ST61Q05	e) Solving an equation like $6x^2 + 5 = 29$	-0.03727	-1.83076	-0.42382
ST61Q07	g) Solving an equation like $2(x+3) = (x + 3)(x - 3)$	0.06218	-1.81880	-0.42771
ST61Q09	i) Solving an equation like $3x+5=17$	-0.02491	-1.87078	-0.35859

Thirteen items measuring students' perceived familiarity with mathematics concepts (*FAMCON*) were used in the Main Survey of PISA 2012 (ST62). Table 16.24 shows the item wording and the international item parameters for this scale as well as the item wording and the item difficulties for three foils that were used to adjust *FAMCON* for overclaiming (or signal detection, see also Chapter 17 on new item formats in PISA 2012). Foils represented non-existing pseudo-concepts and formed the auxiliary scale FOIL. Response categories for students indicating their familiarity with real concepts and with foils were "Never heard of it", "Heard of it once or twice", "Heard of it a few times", "Heard of it often" and "Know it well, understand the concept". If students indicated that they had heard of these pseudo-concepts or even know them well, this would indicate overclaiming. In other words, higher values on *FOIL* were indicative of greater signal detection in terms of students making unsubstantiated claims. An additional index was constructed as *FAMCONC* = *FAMCON* - *FOIL* to adjust the scale indicating familiarity with mathematical concepts for signal detection - or overclaiming.

Table 16.24 Item parameters for familiarity with mathematics concepts (*FAMCON*) and foils used for signal detection adjustment

Item	Thinking about mathematical concepts: how familiar are you with the following terms?	Parameter estimates			
		Delta	tau_1	tau_2	tau_3
ST62Q01	a) Exponential Function	0.70602	0.17846	-0.30139	0.03381
ST62Q02	b) Divisor	-0.39879	0.03620	0.09771	-0.02375
ST62Q03	c) Quadratic Function	-0.09809	0.09546	-0.14735	-0.03443
ST62Q06	e) Linear Equation	-0.18839	0.26520	-0.07690	-0.14734
ST62Q07	f) Vectors	0.24874	0.38656	-0.17589	-0.09203
ST62Q08	g) Complex Number	0.42614	-0.09410	-0.27074	0.09469
ST62Q09	h) Rational Number	-0.28493	0.08260	-0.12672	-0.11466
ST62Q10	i) Radicals	-0.26031	0.18900	0.05347	-0.05948
ST62Q12	k) Polygon	-0.25186	0.37874	-0.09779	-0.09941
ST62Q15	m) Congruent Figure	0.14285	0.45932	-0.19293	0.00194
ST62Q16	n) Cosine	0.12545	0.85226	-0.11527	-0.25220
ST62Q17	o) Arithmetic Mean	0.11169	0.51964	-0.13652	-0.02297
ST62Q19	p) Probability	-0.27852	0.23043	-0.11633	-0.05049
	Foils used for signal detection adjustment				
ST62Q04	d) <Proper Number>	-0.59662	-0.16511	-0.33622	0.16107
ST62Q11	j) <Subjunctive Scaling>	0.35197	0.01429	-0.50321	0.09209
ST62Q13	l) <Declarative Fraction>	0.24465	0.04043	-0.41630	0.06777

Five items measuring teacher behaviour when giving directed instruction (*TCHBEHTD*) were used in the Main Survey of PISA 2012. Table 16.25 shows the item wording and the international item parameters for this scale. Response categories were "Every lesson", "Most lessons", "Some lessons" and "Never or hardly ever". All items were reversed. Item difficulties vary considerably within this index indicating that teachers summarising the previous lesson at the beginning of the next one is reported to occur less frequently than teachers telling students what they have to learn.

Table 16.25 Item parameters for teacher behaviour - teacher directed instruction (*TCHBEHTD*)

Item	How often do these things happen in your mathematics lessons?	Parameter estimates		
		Delta	tau_1	tau_2
ST79Q01	a) The teacher sets clear goals for our learning	-0.19996	-1.24116	-0.00712
ST79Q02	b) The teacher asks me or my classmates to present our thinking or reasoning at some length	0.44775	-1.22218	0.07681
ST79Q06	f) The teacher asks questions to check whether we have understood what was taught	-0.28343	-1.17827	0.08576
ST79Q08	h) At the beginning of a lesson, the teacher presents a short summary of the previous lesson	0.61041	-1.01121	0.28672
ST79Q15	l) The teacher tells us what we have to learn	-0.57477	-1.19271	-0.00089

Four items measuring teacher behaviour when performing student orientation (*TCHBEHSO*) were used in the Main Survey of PISA 2012. Table 16.26 shows the item wording and the international item parameters for this scale. Response categories were "Every lesson", "Most lessons", "Some lessons" and "Never or hardly ever". All items were reversed.

Table 16.26 Item parameters for teacher behaviour - student orientation (*TCHBEHSO*)

Item	How often do these things happen in your mathematics lessons?	Parameter estimates		
		Delta	tau_1	tau_2
ST79Q03	c) The teacher gives different work to classmates who have difficulties learning and/or to those who can advance faster	-0.22534	-0.36141	-0.11517
ST79Q04	d) The teacher assigns projects that require at least one week to complete	0.18333	-0.63919	0.24772
ST79Q07	g) The teacher has us work in small groups to come up with joint solutions to a problem or task	-0.14024	-0.73802	0.22400
ST79Q10	i) The teacher asks us to help plan classroom activities or topics	0.18225	-0.57358	0.10084

Four items measuring teacher behaviour when conducting formative assessment (*TCHBEHFA*) were used in the Main Survey of PISA 2012. Table 16.27 shows the item wording and the international item parameters for this scale. Response categories ranged from "Every lesson" to "Never or hardly ever". All items were reversed.

Table 16.27 Item parameters for teacher behaviour - formative assessment (*TCHBEHFA*)

Item	How often do these things happen in your mathematics lessons?	Parameter estimates		
		Delta	tau_1	tau_2
ST79Q05	e) The teacher tells me about how well I am doing in my mathematics class	0.37671	-1.33131	0.31409
ST79Q11	j) The teacher gives me feedback on my strengths and weaknesses in mathematics	0.57510	-1.14889	0.18644
ST79Q12	k) The teacher tells us what is expected of us when we get a test, quiz or assignment	-0.62686	-1.33841	0.01580
ST79Q17	m) The teacher tells me what I need to do to become better in mathematics	-0.32495	-1.07529	0.11227

Five items measuring teacher support (*TEACHSUP*) in mathematics classes were used in the Main Survey of PISA 2012. Table 16.28 shows the item wording and the international item parameters for this scale. Response categories were "Every lesson", "Most lessons", "Some lessons", "Never or hardly ever". All items were reversed. For this index, item difficulties do not vary considerably which means that some of the listed events were reported to occur only slightly more frequently than others.

Table 16.28 Item parameters for teacher support in mathematics classes (*TEACHSUP*)

Item	How often do these things happen in your mathematics lessons?	Parameter estimates		
		Delta	tau_1	tau_2
ST77Q01	a) The teacher shows an interest in every student's learning	0.21139	-1.84713	0.23901
ST77Q02	b) The teacher gives extra help when students need it	-0.15553	-1.70507	0.20380
ST77Q04	c) The teacher helps students with their learning	-0.23982	-1.51771	0.18839
ST77Q05	d) The teacher continues teaching until the students understand	0.06497	-1.43938	0.21261
ST77Q06	e) The teacher gives students an opportunity to express opinions	0.11899	-1.44889	0.17471

Nine items measuring cognitive activation in mathematics lessons (*COGACT*) were used in the Main Survey of PISA 2012. Table 16.29 shows the item wording and the international item parameters for this scale. Response categories were "Always or almost always", "Often", "Sometimes" and "Never or rarely". All items were reversed, so the higher difficulty corresponds to the lower frequency of the event in the classroom. From the students' points of view, teachers asking students to decide on their own procedures for solving complex problems and teachers presenting problems for which there was no immediately obvious method of solution occurred far less frequently in the classroom than other activities.

Table 16.29 Item parameters for cognitive activation in mathematics lessons (*COGACT*)

Item	Thinking about the mathematics teacher that taught your last mathematics class: How often does each of the following happen?	Parameter estimates		
		Delta	tau_1	tau_2
ST80Q01	a) The teacher asks questions that make us reflect on the problem	-0.11581	-1.64361	0.12254
ST80Q04	b) The teacher gives problems that require us to think for an extended time	0.11324	-1.99125	0.26411
ST80Q05	c) The teacher asks us to decide on our own procedures for solving complex problems	0.61110	-1.34058	0.15393
ST80Q06	d) The teacher presents problems for which there is no immediately obvious method of solution	0.54115	-1.45498	0.12179
ST80Q07	e) The teacher presents problems in different contexts so that students know whether they have understood the concepts	-0.06531	-1.55612	0.09338
ST80Q08	f) The teacher helps us to learn from mistakes we have made	-0.18099	-1.20016	0.03615
ST80Q09	g) The teacher asks us to explain how we have solved a problem	-0.43909	-1.28211	0.08367
ST80Q10	h) The teacher presents problems that require students to apply what they have learned to new contexts	-0.17076	-1.48658	0.10387
ST80Q11	i) The teacher gives problems that can be solved in several different ways	-0.29353	-1.75602	0.17498

This scale provides information on mathematics teacher support (*MTSUP*). There are four items in this scale. The four response categories vary from “Strongly agree” to “Strongly disagree”. All items were reversed. From Table 16.30 which shows the item wording and the international item parameters for this scale it can be seen that students find it harder to agree with the statement that teachers give them the opportunity to express opinions.

Table 16.30 Item parameters mathematics teacher support (*MTSUP*)

Item	Thinking about the mathematics teacher who taught your last mathematics class: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST83Q01	a) My teacher lets us know we need to work hard	-0.27208	-1.97628	-0.79466
ST83Q02	b) My teacher provides extra help when needed	0.02497	-1.70752	-0.75517
ST83Q03	c) My teacher helps students with their learning	-0.08478	-1.72847	-0.80214
ST83Q04	d) My teacher gives students the opportunity to express opinions	0.33189	-1.64770	-0.70635

This scale provides information on classroom management (*CLSMAN*) and consists of four items. The four response categories vary from “Strongly agree” to “Strongly disagree”. Table 16.31 shows the item wording and the international item parameters for this scale. All items except the last one (ST85Q04) were reversed.

Table 16.31 Item parameters classroom management (*CLSMAN*)

Item	Thinking about the mathematics teacher who taught your last mathematics class: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
ST85Q01	a) My teacher gets students to listen to him or her	-0.27661	-1.46872	-0.60820
ST85Q02	b) My teacher keeps the class orderly	-0.14614	-1.66419	-0.44530
ST85Q03	c) My teacher starts lessons on time	-0.25532	-1.57764	-0.34718
ST85Q04	d) The teacher has to wait a long time for students to <quiet down>	0.67807	-1.49277	-0.21922

This scale provides information on disciplinary climate in the classroom (*DISCLIMA*) based on five items. The four response categories were “Every lesson”, “Most lessons”, “Some lessons”, to “Never or hardly ever”. Table 16.32 shows the item wording and the international item parameters for this scale.

Table 16.32 Item parameters for disciplinary climate (*DISCLIMA*)

Item	How often do these things happen in your mathematics lessons?	Parameter estimates		
		Delta	tau_1	tau_2
ST81Q01	a) Students don’t listen to what the teacher says	0.35916	-2.06346	-0.61398
ST81Q02	b) There is noise and disorder	0.19734	-1.73779	-0.50655
ST81Q03	c) The teacher has to wait a long time for students to <quiet down>	-0.09943	-1.54240	-0.41861
ST81Q04	d) Students cannot work well	-0.30906	-1.71494	-0.55403
ST81Q05	e) Students don’t start working for a long time after the lesson begins	-0.14801	-1.44841	-0.40606

School climate

In PISA 2012, school climate was covered by two scaled indices based on responses to the Student Questionnaire as listed in Table 16.33.

Table 16.33 School climate indices

Index	Index label	Relationship to other PISA surveys
STUDREL	Teacher-Student Relation	Used in 2000, 2003 and 2009; STUREL in 2003; not used in 2006.
BELONG	Sense of Belonging to School	Similar index was used in 2000 and 2003, but the scale included only 6 items with the wording similar to the first 6 items of 2012.The stem and items 2, 3, 4 and 6 have been modified from the 2003 Student Questionnaire and three more items were added. Not used in 2006 and 2009.

Table 16.34 shows high degree of internal consistency across OECD countries for teacher-student relations (*STUDREL*) and moderate to low reliability for sense of belonging to school (*BELONG*).

Table 16.34 Scale reliabilities for school climate indices in OECD countries

	STUDREL	BELONG
Australia	0.85	0.32
Austria	0.82	0.39
Belgium	0.80	0.38
Canada	0.85	0.35
Chile	0.83	0.44
Czech Republic	0.83	0.39
Denmark	0.82	0.36
Estonia	0.81	0.40
Finland	0.83	0.31
France	0.78	0.32
Germany	0.81	0.38
Greece	0.80	0.47
Hungary	0.83	0.44
Iceland	0.88	0.45
Ireland	0.83	0.27
Israel	0.85	0.43
Italy	0.81	0.41
Japan	0.86	0.35
Korea	0.83	0.43
Luxembourg	0.83	0.48
Mexico	0.81	0.53
Netherlands	0.78	0.44
New Zealand	0.84	0.35
Norway	0.86	0.38
Poland	0.84	0.45
Portugal	0.84	0.49
Slovak Republic	0.80	0.43
Slovenia	0.79	0.49
Spain	0.82	0.46
Sweden	0.86	0.52
Switzerland	0.82	0.39
Turkey	0.76	0.47
United Kingdom	0.85	0.31
United States	0.83	0.32
OECD median	**0.83**	**0.40**

Similar to OECD countries Table 16.35 shows a high degree of internal consistency across partner countries and economies for teacher-student relations (*STUDREL*) and moderate to low reliability for sense of belonging to school (*BELONG*).

Table 16.35 Scale reliabilities for school climate indices in partner countries and economies

	STUDREL	BELONG
Albania	0.76	0.47
Argentina	0.80	0.60
Brazil	0.81	0.59
Bulgaria	0.82	0.65
Colombia	0.82	0.57
Costa Rica	0.79	0.49
Croatia	0.84	0.39
Cyprus[1, 2]	0.83	0.57
Hong Kong-China	0.84	0.41
Indonesia	0.71	0.55
Jordan	0.78	0.65
Kazakhstan	0.83	0.53
Latvia	0.79	0.32
Liechtenstein	0.83	0.43
Lithuania	0.82	0.43
Macao-China	0.83	0.35
Malaysia	0.73	0.53
Montenegro	0.85	0.62
Peru	0.80	0.60
Qatar	0.79	0.67
Romania	0.76	0.55
Russian Federation	0.81	0.55
Serbia	0.83	0.53
Shanghai-China	0.87	0.35
Singapore	0.83	0.32
Chinese Taipei	0.86	0.36
Thailand	0.83	0.58
Tunisia	0.76	0.51
United Arab Emirates	0.81	0.51
Uruguay	0.79	0.57
Viet Nam	0.75	0.45
Median	**0.81**	**0.53**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Five items on teacher-student relations were included in the Student Questionnaire. This scale provides information on students' perceived teacher's interest in student performance. There are four response categories varying from "Strongly agree", "Agree", "Disagree" to "Strongly disagree". All items were reversed. Table 16.36 shows the item wording and the international item parameters for this scale. The statement that students found the most difficult to agree with was that most of their teachers really listened to what students had to say.

Table 16.36 Item parameters for teacher-student relations (*STUDREL*)

		Parameter estimates		
Item	Thinking about the teachers at your school: to what extent do you agree with the following statements?	Delta	tau_1	tau_2
ST86Q01	a) Students get along well with most teachers	-0.25440	-2.61793	-0.67095
ST86Q02	b) Most teachers are interested in students' well-being	0.02701	-2.58166	-0.58553
ST86Q03	c) Most of my teachers really listen to what I have to say	0.26661	-2.63891	-0.47193
ST86Q04	d) If I need extra help, I will receive it from my teachers	-0.11302	-2.23629	-0.77270
ST86Q05	e) Most of my teachers treat me fairly	0.07380	-2.02169	-0.91774

Nine items on sense of belonging to school were included in the Student Questionnaire. There were four response categories varying from "Strongly agree", "Agree", "Disagree" to "Strongly disagree". All items except *ST87Q01, ST87Q04* and *ST87Q06* were reversed. Table 16.37 shows the item wording and the international item parameters for this scale. As could expected among 15-year-olds, the statement that students found the most difficult to endorse was the statement that things were ideal in their school.

Table 16.37 **Item parameters for sense of belonging to school (*BELONG*)**

Item	Thinking about your school: to what extent do you agree with the following statements?	Parameter estimates Delta	tau_1	tau_2
ST87Q01	a) I feel like an outsider (or left out of things) at school	-0.19030	-0.74122	-0.88944
ST87Q02	b) I make friends easily at school	-0.33145	-1.54867	-0.81858
ST87Q03	c) I feel like I belong at school	0.04298	-1.40019	-0.73306
ST87Q04	d) I feel awkward and out of place in my school	-0.16960	-0.98328	-0.68662
ST87Q05	e) Other students seem to like me	-0.01611	-1.63572	-1.07022
ST87Q06	f) I feel lonely at school	-0.38428	-0.74253	-0.79140
ST87Q07	g) I feel happy at school	0.09349	-1.40815	-0.79582
ST87Q08	h) Things are ideal in my school	0.73038	-1.83857	-0.26238
ST87Q09	i) I am satisfied with my school	0.22489	-1.24562	-0.79514

Attitudes towards School

In PISA 2012, the attitudes towards school was covered by two scaled indices based on eight items in the Student Questionnaire (ST88, ST89) as listed in Table 16.38.

Table 16.38 **Attitude towards school indices**

Index	Index label	Relationship to other PISA surveys
ATSCHL	Attitude towards School: Learning Outcomes	New
ATTLNACT	Attitude towards School: Learning Activities	New

Table 16.39 shows a high degree of internal consistency across OECD countries for attitude towards school in terms of learning activities (*ATTLNACT*) and moderate to high reliability for attitude towards school regarding learning outcomes (*ATSCHL*).

Table 16.39 **Scale reliabilities and correlations for attitude towards school indices in OECD countries**

	ATSCHL	ATTLNACT	Correlation
Australia	0.73	0.87	0.52
Austria	0.73	0.68	0.43
Belgium	0.65	0.77	0.40
Canada	0.74	0.85	0.50
Chile	0.65	0.79	0.38
Czech Republic	0.65	0.76	0.43
Denmark	0.68	0.81	0.40
Estonia	0.71	0.80	0.46
Finland	0.74	0.82	0.55
France	0.70	0.78	0.46
Germany	0.67	0.67	0.38
Greece	0.67	0.69	0.48
Hungary	0.65	0.78	0.50
Iceland	0.68	0.87	0.48
Ireland	0.76	0.85	0.48
Israel	0.73	0.80	0.44
Italy	0.69	0.74	0.48
Japan	0.70	0.78	0.45
Korea	0.74	0.81	0.44
Luxembourg	0.62	0.78	0.44
Mexico	0.55	0.78	0.43
Netherlands	0.55	0.81	0.42
New Zealand	0.72	0.86	0.48
Norway	0.72	0.83	0.44
Poland	0.71	0.79	0.50
Portugal	0.71	0.85	0.50
Slovak Republic	0.64	0.75	0.47
Slovenia	0.61	0.78	0.46
Spain	0.69	0.80	0.47
Sweden	0.68	0.85	0.46
Switzerland	0.67	0.72	0.40
Turkey	0.59	0.80	0.45
United Kingdom	0.73	0.86	0.45
United States	0.75	0.87	0.50
OECD median α	**0.69**	**0.80**	
OECD avarage correlation			**0.46**
Correlation S.D.			**0.04**

The internal consistency of the scale indicating attitudes towards school in terms of learning activities (*ATTLNACT*) was also high in partner countries and economies (see Table 16.40) and moderate to high for the scale indicating attitudes towards school regarding learning outcomes (*ATSCHL*). Correlations for these two scales were positive and consistent across both OECD and partner countries and economies as indicated by the low standard deviation of the correlation (0.04). This supports the construct validity of these scales across all participating countries.

Table 16.40 Scale reliabilities and correlations for attitude towards school indices in partner countries and economies

	ATSCHL	ATTLNACT	Correlation
Albania	0.54	0.75	0.44
Argentina	0.56	0.79	0.43
Brazil	0.58	0.81	0.44
Bulgaria	0.52	0.82	0.48
Colombia	0.62	0.77	0.44
Costa Rica	0.59	0.81	0.43
Croatia	0.68	0.80	0.47
Cyprus[1, 2]	0.63	0.82	0.53
Hong Kong-China	0.68	0.82	0.43
Indonesia	0.60	0.82	0.44
Jordan	0.29	0.77	0.43
Kazakhstan	0.60	0.84	0.48
Liechtenstein	0.60	0.74	0.45
Lithuania	0.67	0.72	0.52
Latvia	0.66	0.78	0.45
Macao-China	0.62	0.79	0.41
Malaysia	0.51	0.83	0.51
Montenegro	0.59	0.79	0.44
Peru	0.52	0.77	0.42
Qatar	0.45	0.83	0.44
Romania	0.57	0.75	0.48
Russian Federation	0.70	0.82	0.48
Serbia	0.41	0.75	0.41
Shanghai-China	0.75	0.78	0.43
Singapore	0.62	0.81	0.38
Chinese Taipei	0.70	0.84	0.48
Thailand	0.43	0.79	0.40
Tunisia	0.63	0.78	0.51
United Arab Emirates	0.59	0.80	0.45
Uruguay	0.55	0.80	0.37
Viet Nam	0.65	0.61	0.36
Median	**0.60**	**0.79**	
Avarage correlation			**0.45**
Correlation SD			**0.04**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Four items were included in the attitude towards school regarding learning outcomes (*ATSCHL*) scale. All four items had four response categories from "Strongly agree", "agree", "Disagree" to "Strongly disagree" and items *ST88Q03* and *ST88Q04* were reversed. Table 16.41 shows the item wording and the international item parameters for this scale. Results show that students found it hardest to strongly disagree with the statement that school had done little to prepare them for adult life after leaving school.

Table 16.41 Item parameters for attitude towards school: Learning outcomes (*ATSCHL*)

		Parameter estimates		
Item	Thinking about what you have learned at school: to what extent do you agree with the following statements?	Delta	tau_1	tau_2
ST88Q01	a) School has done little to prepare me for adult life when I leave school	0.69821	-1.45109	-0.55021
ST88Q02	b) School has been a waste of time	-0.25285	-0.68621	-1.08584
ST88Q03	c) School has helped give me confidence to make decisions	-0.01432	-1.59410	-0.66939
ST88Q04	d) School has taught me things which could be useful in a job	-0.43104	-0.99365	-0.89887

Four items regarding attitude towards school in terms of learning activities (*ATTLNACT*) were included in the Student Questionnaire with four response categories from "Strongly agree" to "Strongly disagree". All items were reversed. Table 16.42 shows the item wording and the international item parameters for this scale, the latter indicating that difficulty levels were similar for these four attitudinal items.

Table 16.42 Item parameters for attitude towards school: Learning activities (*ATTLNACT*)

		Parameter estimates		
Item	Thinking about your school: to what extent do you agree with the following statements?	Delta	tau_1	tau_2
ST89Q02	a) Trying hard at school will help me get a good job	0.20147	-2.11171	-0.89032
ST89Q03	b) Trying hard at school will help me get into a good <college>	-0.04516	-1.75930	-1.05927
ST89Q04	c) I enjoy receiving good <grades>	-0.28492	-1.70583	-1.08044
ST89Q05	d) Trying hard at school is important	0.12861	-1.88420	-1.10550

Problem Solving

In PISA 2012, the two new scaled indices (see Table 16.43), namely perseverance and openness to problem solving were developed in recognition of the increasing importance of problems solving in the cognitive part of the assessment.

Table 16.43 Problem solving indices

Index	Index label	Relationship to other PISA surveys
PERSEV	Perseverance	New
OPENPS	Openness for Problem Solving	New

Tables 16.44 and 16.45 show a high degree of internal consistency for both OECD and partner countries and economies for the two problem solving indices.

Table 16.44 Scale reliabilities for problem solving indices in OECD countries

	PERSEV	OPENPS
Australia	0.87	0.84
Austria	0.68	0.80
Belgium	0.77	0.81
Canada	0.85	0.85
Chile	0.79	0.80
Czech Republic	0.76	0.80
Denmark	0.81	0.83
Estonia	0.80	0.84
Finland	0.82	0.85
France	0.78	0.83
Germany	0.67	0.81
Greece	0.69	0.77
Hungary	0.78	0.81
Iceland	0.87	0.89
Ireland	0.85	0.81
Israel	0.80	0.80
Italy	0.74	0.78
Japan	0.78	0.83
Korea	0.81	0.81
Luxembourg	0.78	0.83
Mexico	0.78	0.84
Netherlands	0.81	0.83
New Zealand	0.86	0.84
Norway	0.83	0.88
Poland	0.79	0.86
Portugal	0.85	0.84
Slovak Republic	0.75	0.80
Slovenia	0.78	0.80
Spain	0.80	0.80
Sweden	0.85	0.86
Switzerland	0.72	0.82
Turkey	0.80	0.78
United Kingdom	0.86	0.82
United States	0.87	0.85
OECD median	**0.80**	**0.82**

Table 16.45 Scale reliabilities for problem solving indices in partner countries and economies

	PERSEV	OPENPS
Albania	0.75	0.76
Argentina	0.79	0.80
Brazil	0.81	0.81
Bulgaria	0.82	0.81
Colombia	0.77	0.79
Costa Rica	0.81	0.81
Croatia	0.80	0.74
Cyprus[1, 2]	0.82	0.81
Hong Kong-China	0.82	0.86
Indonesia	0.82	0.81
Jordan	0.77	0.80
Kazakhstan	0.84	0.83
Latvia	0.78	0.80
Liechtenstein	0.74	0.82
Lithuania	0.72	0.80
Macao-China	0.79	0.82
Malaysia	0.83	0.81
Montenegro	0.79	0.74
Peru	0.77	0.78
Qatar	0.83	0.81
Romania	0.75	0.77
Russian Federation	0.82	0.81
Serbia	0.75	0.80
Shanghai-China	0.78	0.84
Singapore	0.81	0.81
Chinese Taipei	0.84	0.86
Thailand	0.79	0.82
Tunisia	0.78	0.75
United Arab Emirates	0.80	0.78
Uruguay	0.80	0.80
Viet Nam	0.61	0.80
Median	**0.79**	**0.81**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Five items measuring perseverance (*PERSEV*) were included in the Student Questionnaire which had five response categories, namely "Very much like me", "Mostly like me", "Somewhat like me", "Not much like me" and "Not at all like me". The last three items were reversed. Table 16.46 shows the item wording and the international item parameters for this scale.

Table 16.46 Item parameters for perseverance (*PERSEV*)

		Parameter estimates			
Item	How well does each of the following statements below describe you?	Delta	tau_1	tau_2	tau_3
ST93Q01	a) When confronted with a problem, I give up easily	-0.05108	-0.58760	-0.69646	0.08203
ST93Q03	b) I put off difficult problems	0.28741	-0.91697	-0.62301	0.40159
ST93Q04	c) I remain interested in the tasks that I start	-0.20042	-1.19575	-0.51577	0.38405
ST93Q06	d) I continue working on tasks until everything is perfect	-0.11046	-1.23633	-0.33077	0.49403
ST93Q07	e) When confronted with a problem, I do more than what is expected of me	0.07455	-1.22383	-0.37935	0.57732

Five items on openness for problem solving (*OPENPS*) were included in the Student Questionnaire with five response categories, namely "Very much like me", "Mostly like me", "Somewhat like me", "Not much like me" and "Not at all like me". All items were reversed. Table 16.47 shows the item wording and the international item parameters for this scale.

Table 16.47 Item parameters for openness for problem solving (*OPENPS*)

		Parameter estimates			
Item	How well does each of the following statements below describe you?	Delta	tau_1	tau_2	tau_3
ST94Q05	a) I can handle a lot of information	-0.08964	-2.30286	-0.90555	0.83950
ST94Q06	b) I am quick to understand things	-0.20682	-2.21470	-0.89174	0.68774
ST94Q09	c) I seek explanations for things	-0.35650	-2.09394	-0.88564	0.75116
ST94Q10	d) I can easily link facts together	-0.17628	-2.27798	-0.77326	0.76166
ST94Q14	e) I like to solve complex problems	0.82924	-1.61131	-0.51812	0.61360

ICT familiarity

The ICT familiarity questionnaire was an optional instrument, which was administered in 42 of the participating countries in PISA 2012. Eight scaled indices (see Table 16.48) were computed based on the information obtained from this questionnaire.

Table 16.48 ICT familiarity indices

Index	Index label	Relationship to other PISA surveys
ICTHOME	ICT Availability at Home	Modified from 2009, new items added
ICTSCH	ICT Availability at School	Modified from 2009, new items added
ENTUSE	ICT Entertainment Use	Modified from 2009, new items added
HOMSCH	ICT Use at Home for School-related Tasks	Modified from 2009
USESCH	Use of ICT at School	Modified from 2009
USEMATH	Use of ICT in Mathematics Lessons	New
ICTATTPOS	Attitudes Towards Computers: Computer as a Tool for School Learning	New
ICTATTNEG	Attitudes Towards Computers: Limitations of the Computer as a Tool for School Learning	New

Table 16.49 shows various degree of internal consistency across OECD countries for ICT familiarity indices. The OECD median for Cronbach's alpha varied from medium for ICT availability at home (*ICTHOME*, median α=0.53) to very high for the use of ICT in mathematics lessons (*USEMATH*, median α=0.91). These results were similar for partner countries and economies as can be seen in Table 16.50.

Table 16.49 Scale reliabilities for ICT familiarity indices in OECD countries

	ICTHOME	ICTSCH	ENTUSE	HOMSCH	USESCH	USEMATH	ICTATTPOS	ICTATTNEG
Australia	0.53	0.40	0.78	0.87	0.78	0.89	0.76	0.74
Austria	0.44	0.60	0.77	0.82	0.81	0.87	0.74	0.72
Belgium	0.50	0.70	0.77	0.85	0.88	0.89	0.76	0.66
Chile	0.68	0.65	0.83	0.83	0.84	0.91	0.76	0.65
Czech Republic	0.45	0.55	0.75	0.81	0.85	0.90	0.68	0.65
Denmark	0.49	0.45	0.74	0.80	0.80	0.91	0.77	0.78
Estonia	0.45	0.63	0.70	0.75	0.84	0.88	0.74	0.71
Finland	0.41	0.53	0.73	0.81	0.81	0.89	0.78	0.67
Germany	0.42	0.66	0.75	0.75	0.82	0.88	0.75	0.73
Greece	0.64	0.68	0.86	0.88	0.91	0.92	0.67	0.65
Hungary	0.64	0.65	0.80	0.84	0.88	0.92	0.76	0.71
Iceland	0.40	0.56	0.74	0.85	0.87	0.92	0.80	0.73
Ireland	0.50	0.67	0.78	0.80	0.81	0.91	0.75	0.74
Israel	0.65	0.76	0.80	0.84	0.91	0.94	0.72	0.68
Italy	0.56	0.72	0.80	0.82	0.87	0.89	0.68	0.71
Japan	0.63	0.71	0.78	0.70	0.74	0.96	0.81	0.71
Korea	0.55	0.68	0.74	0.83	0.88	0.95	0.81	0.70
Mexico	0.81	0.68	0.88	0.86	0.87	0.93	0.77	0.69
Netherlands	0.51	0.49	0.70	0.75	0.83	0.91	0.75	0.65
New Zealand	0.60	0.52	0.80	0.88	0.83	0.91	0.80	0.81
Norway	0.50	0.38	0.75	0.83	0.84	0.84	0.79	0.79
Poland	0.59	0.68	0.79	0.80	0.89	0.92	0.73	0.73
Portugal	0.55	0.62	0.83	0.88	0.92	0.93	0.68	0.78
Slovak Republic	0.61	0.66	0.83	0.84	0.87	0.92	0.70	0.67
Slovenia	0.56	0.69	0.81	0.86	0.92	0.92	0.75	0.71
Spain	0.51	0.61	0.76	0.83	0.83	0.92	0.69	0.68
Sweden	0.53	0.47	0.77	0.87	0.88	0.94	0.82	0.79
Switzerland	0.47	0.66	0.77	0.83	0.85	0.89	0.75	0.72
Turkey	0.78	0.75	0.90	0.86	0.89	0.92	0.79	0.77
OECD median	**0.53**	**0.65**	**0.78**	**0.83**	**0.85**	**0.91**	**0.75**	**0.71**

Table 16.50 Scale reliabilities for ICT familiarity indices in partner countries and economies

	ICTHOME	ICTSCH	ENTUSE	HOMSCH	USESCH	USEMATH	ICTATTPOS	ICTATTNEG
Costa Rica	0.76	0.69	0.87	0.83	0.86	0.94	0.75	0.63
Croatia	0.52	0.67	0.81	0.81	0.87	0.92	0.73	0.69
Hong Kong-China	0.52	0.55	0.74	0.82	0.85	0.90	0.67	0.61
Jordan	0.85	0.79	0.93	0.90	0.93	0.93	0.72	0.74
Latvia	0.57	0.67	0.79	0.82	0.89	0.92	0.73	0.69
Liechtenstein	0.38	0.65	0.78	0.83	0.83	0.85	0.79	0.73
Macao-China	0.56	0.61	0.79	0.80	0.81	0.90	0.74	0.72
Russian Federation	0.60	0.69	0.84	0.85	0.91	0.92	0.73	0.70
Serbia	0.70	0.74	0.84	0.85	0.89	0.93	0.78	0.75
Shanghai-China	0.65	0.70	0.86	0.81	0.82	0.91	0.72	0.71
Singapore	0.53	0.63	0.79	0.89	0.87	0.91	0.73	0.72
Chinese Taipei	0.63	0.59	0.83	0.86	0.85	0.95	0.81	0.72
Uruguay	0.74	0.72	0.84	0.84	0.90	0.92	0.77	0.71
Median	**0.60**	**0.67**	**0.83**	**0.83**	**0.87**	**0.92**	**0.73**	**0.71**

Eleven items provided information on ICT availability at home (*ICTHOME*) in the ICT familiarity questionnaire in PISA 2012. The three response categories were "Yes, and I use it", "Yes, but I don't use it" and "No". All items were reversed. Table 16.51 shows the devices for which availability and use at home were checked as well as the international IRT parameters for this scale. The distribution of item difficulties and step difficulties for this scale are reasonable and appropriate with tablets and ebook readers not as often used as desktop computers or cell phones.

Table 16.51 Item parameters for ICT availability at home (*ICTHOME*)

		Parameter estimates	
Item	Are any of these devices available for you to use at home?	Delta	tau_1
IC01Q01	a) Desktop computer	-0.37724	0.90409
IC01Q02	b) Portable laptop, or notebook	-0.07181	1.26122
IC01Q03	c) <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook ™>)	1.21142	1.56453
IC01Q04	d) Internet connection	-1.00973	2.17172
IC01Q05	e) <Video games console>, e.g. <Sony® PlayStation® >	0.35765	0.88705
IC01Q06	f) <Cell phone> (without Internet access)	0.10868	0.56614
IC01Q07	g) <Cell phone> (with Internet access)	-0.30415	1.15461
IC01Q08	h) Portable music player (Mp3/Mp4 player, iPod® or similar)	-0.35090	0.99946
IC01Q09	i) Printer	-0.18337	1.02506
IC01Q10	j) USB (memory) stick	-1.00313	0.72701
IC01Q11	k) <eBook reader>, e.g. <Amazon® Kindle ™>	1.62258	0.95641

Seven items provided information on ICT availability at school (*ICTSCH*) in PISA 2012. Again, the three response categories included "Yes, and I use it", "Yes, but I don't use it" and "No". All items were reversed. Table 16.52 shows the devices for which availability and use at school were checked as well as the international IRT parameters for this scale. The distribution of item difficulties and step difficulties for this scale are reasonable and appropriate with tablets and e-book readers not as often used at school as desktop computers or internet connections.

Table 16.52 Item parameters for ICT availability at school (*ICTSCH*)

		Parameter estimates	
Item	Are any of these devices available for you to use at school?	Delta	tau_1
IC02Q01	a) Desktop computer	-1.48349	-0.18181
IC02Q02	b) Portable laptop or notebook	0.43582	0.45335
IC02Q03	c) <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook ™>)	1.70674	0.92454
IC02Q04	d) Internet connection	-1.53021	-0.06362
IC02Q05	e) Printer	-0.99947	-0.24697
IC02Q06	f) USB (memory) stick	0.14648	0.28204
IC02Q07	g) <eBook reader>, e.g. <Amazon® Kindle ™>	1.72413	0.45042

Ten items sought information on use of ICT for entertainment (*ENTUSE*). Five response categories included "Never or hardly ever", "Once or twice a month", "Once or twice a week", "Almost every day" and "Every day". Table 16.53 shows the item wording and international IRT parameters for this scale. The distribution of item and step difficulties for this scale were reasonable and appropriate and indicated that uploading student's own created contents for sharing occurred less often than browsing the Internet for fun.

Table 16.53 Item parameters for ICT entertainment use (*ENTUSE*)

Item	How often do you use a computer for the following activities outside of school?	Parameter estimates Delta	tau_1	tau_2	tau_3
IC08Q01	a) Playing one-player games	0.56025	-0.07895	-0.55524	0.20545
IC08Q02	b) Playing collaborative online games	0.53815	0.48298	-0.48203	-0.02671
IC08Q03	c) Using email	0.05040	-0.50015	-0.37791	0.33186
IC08Q04	d) <Chatting online> (e.g.<MSN®>)	-0.18749	0.52895	-0.47947	-0.14916
IC08Q05	e) Participating in social networks (e.g.<Facebook>, <MySpace>)	-0.61577	0.82483	-0.51794	-0.17790
IC08Q06	f) Browsing the Internet for fun (such as watching videos, e.g. <YouTube™>)	-0.72756	0.00021	-0.60131	0.18488
IC08Q07	g) Reading news on the Internet (e.g.current affairs)	0.00238	-0.16180	-0.53502	0.25195
IC08Q08	h) Obtaining practical information from the Internet (e.g.locations, dates of events)	0.01770	-0.63263	-0.61047	0.48391
IC08Q09	i) Downloading music, films, games or software from the Internet	-0.25440	-0.38493	-0.37840	0.30821
IC08Q11	j) Uploading your own created contents for sharing (e.g.music, poetry, videos, computer programs)	0.61634	0.23946	-0.25311	0.09636

Seven items sought information on the use of ICT outside school but for school related tasks (*HOMSCH*). The five response categories were "Never or hardly ever", "Once or twice a month", "Once or twice a week", "Almost every day" and "Every day". Table 16.54 shows the item wording and international IRT parameters for this scale.

Table 16.54 Item parameters for ICT use at home for school related tasks (*HOMSCH*)

Item	How often do you use a computer for the following activities outside of school?	Parameter estimates Delta	tau_1	tau_2	tau_3
IC09Q01	a) Browsing the Internet for schoolwork (e.g. for preparing an essay or presentation)	-0.47218	-1.84411	-0.61722	0.98683
IC09Q02	b) Using email for communication with other students about schoolwork	-0.07103	-0.63527	-0.62102	0.39880
IC09Q03	c) Using email for communication with teachers and submission of homework or other schoolwork	0.53073	-0.77744	-0.35176	0.56909
IC09Q04	d) Downloading, upload or browse material from my school's website (e.g. time table or course materials)	0.12747	-0.67462	-0.36316	0.36557
IC09Q05	e) Checking the school's website for announcements, e.g. absence of teachers	0.13397	-0.29759	-0.35088	0.16093
IC09Q06	f) Doing homework on the computer	-0.30237	-1.09545	-0.59187	0.57559
IC09Q07	g) Sharing school related materials with other students	0.05341	-0.67695	-0.60679	0.38437

Nine items invited students to report on the use of computers for ICT related activities at school (*USESCH*). Five response categories vary from "Never or hardly ever", "Once or twice a month", "Once or twice a week", "Almost every day" to "Every day". Table 16.55 shows the item wording and international IRT parameters for this scale.

Table 16.55 Item parameters for use of ICT at school (*USESCH*)

Item	How often do you use a computer for the following activities at school?	Parameter estimates Delta	tau_1	tau_2	tau_3
IC10Q01	a) <Chatting on line> at school	-0.04326	0.24104	-0.86815	0.45464
IC10Q02	b) Using email at school	-0.03091	-0.47426	-0.67827	0.65702
IC10Q03	c) Browsing the Internet for schoolwork	-0.66296	-1.16175	-0.75605	0.89120
IC10Q04	d) Downloading, uploading or browsing material from the school's website (e.g. <intranet>)	0.05209	-0.32018	-0.76602	0.47839
IC10Q05	e) Posting my work on the school's website	0.36697	0.01575	-0.83393	0.42845
IC10Q06	f) Playing simulations at school	0.39406	0.07572	-0.67624	0.32083
IC10Q07	g) Practicing and drilling, such as for foreign language learning or mathematics	0.11228	-0.70404	-0.59797	0.66411
IC10Q08	h) Doing homework on a school computer	-0.06137	-0.56136	-0.63843	0.54638
IC10Q09	i) Using school computers for group work and communication with other students	-0.12690	-0.94594	-0.50315	0.75345

A new scale was created in PISA 2012 based on seven items as an indicator of students' reported use of ICT in mathematics lessons (*USEMATH*). The three response categories included "Yes, students did this", "Yes, but only the teacher demonstrated this" to "No". All items were reversed. Table 16.56 shows the item wording and international IRT parameters for this scale.

Table 16.56 Item parameters for use of ICT in mathematics lessons (*USEMATH*)

Item	Within the last month, has a computer ever been used for the following purposes in your mathematics lessons?	Parameter estimates Delta	tau_1
IC11Q01	a) Drawing the graph of a function (such as y = 4x+6)	-0.18024	-0.22941
IC11Q02	b) Calculating with numbers (such as calculating 5*233/8)	0.02833	0.09262
IC11Q03	c) Constructing geometric figures (e.g. an equilateraltriangle with given side lengths)	-0.03530	-0.25815
IC11Q04	d) Entering data in a spreadsheet (e.g. in <Excel ™>)	-0.33025	0.04599
IC11Q05	e) Rewriting algebraic expressions and solving equations (such as $a^2+2ab+b^2$)	0.10770	-0.08542
IC11Q06	f) Drawing histograms (a graph that shows the distribution of frequencies of data)	0.21583	-0.21162
IC11Q07	g) Finding out how the graph of a function like $y=ax^2$ changes depending on a	0.19393	-0.31561

Three items in the ICT familiarity questionnaire formed a scale indicating attitudes towards the computer as a tool for school learning (*ICTATTPOS*). The items had for response categories, namely “Strongly agree”, “Agree”, “Disagree” and “Strongly disagree”. All items were reversed. Table 16.57 shows the item wording and the international item parameters for this scale. As can be seen, students found it harder to endorse the statement that “Doing my homework using a computer makes it more fun” than agreeing with the statement that the Internet being a great resource for students’ school work.

Table 16.57 Item parameters for attitudes towards computers: computer as a tool for school learning (*ICTATTPOS*)

Item	Thinking about your experience with computers: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
IC22Q01	a) The computer is a very useful tool for my schoolwork	-0.24361	-1.95350	-0.81100
IC22Q02	b) Doing my homework using a computer makes it more fun	0.79826	-2.48214	-0.28989
IC22Q04	c) The Internet is a great resource for obtaining information I can use for my school work	-0.55465	-1.58778	-1.01963

Three items on attitudes towards computers limitations of the computer as a tool for school learning (*ICTATTNEG*) were included in the ICT familiarity questionnaire. There are four response categories varying from “Strongly agree”, “Agree”, “Disagree” to “Strongly disagree”. All items were reversed. Table 16.58 shows the item wording and the international item parameters for this scale.

Table 16.58 Item parameters for attitudes towards computers: limitations of the computer as a tool for school learning (*ICTATTNEG*)

Item	Thinking about your experience with computers: to what extent do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
IC22Q06	d) Using the computer for learning is troublesome	0.28503	-2.04715	0.59615
IC22Q07	e) Since anyone can upload information to the internet, it is in general not suitable to use it for schoolwork	-0.22365	-2.43535	0.47849
IC22Q08	f) Information obtained from the internet is generally too unreliable to be used for school assignments	-0.06138	-2.47303	0.62212

Educational career

The educational career questionnaire was an optional instrument, some parts of which were administered in 23 of the countries/economies participating in PISA 2012. Of these, 22 countries/economies collected sufficient information to compute some of the six scaled indices listed in Table 16.59.

It should be noted that not all 22 countries/economies administered all questions so that the information in the following tables is provided for only those countries/economies that had administered the questions forming a scale.

Table 16.59 Educational career indices

Index	Index label	Relationship to other PISA surveys
INFOCAR	Information about Careers	New
INFOJOB1	Information about the Labour Market provided by the School	New
INFOJOB2	Information about the Labour Market provided outside of School	New
HOSTCUL	Acculturation: Host Culture Oriented Strategies	New
HERITCUL	Acculturation: Heritage Culture Oriented Strategies	New
CULTDIST	Cultural Distance between Host and Heritage Culture	New

Table 16.60 shows an acceptable level of internal consistency across those OECD countries that administered questions for the educational career indices with only the scale indicating students’ reported activities to obtain career information having a slightly lower reliability.

Table 16.60 Scale reliabilities for educational career indices in OECD countries

	INFOCAR	INFOJOB1	INFOJOB2	HOSTCUL	HERITCUL	CULTDIST
Australia	0.67	0.79	0.76	*	*	*
Austria	0.59**	0.72	0.68	*	*	*
Belgium	0.60	0.62	0.64	0.85	0.83	0.80
Canada	0.64	0.79	0.75	*	*	*
Denmark	0.53	0.61	0.62	0.88	0.81	0.75
Finland	0.67**	0.79	0.77	0.86	0.84	0.79
Germany	*	*	*	*	*	*
Hungary	0.64	0.67	0.70	*	*	*
Ireland	0.70	0.79	0.73	*	*	*
Italy	0.59	0.63	0.58	0.82	0.81	0.79
Korea	0.67	0.78	0.79	0.92	0.94	0.95
Luxembourg	0.69	0.78	0.73	*	*	*
New Zealand	0.65**	0.76	0.75	*	*	*
Portugal	0.69	0.80	0.79	0.86	0.84	0.82
Slovak Republic	0.71	0.71	0.71	0.85	0.85	0.80
Slovenia	0.77	0.75	0.73	0.81	0.84	0.78

* All questions from the index deleted by the country.
** Item EC03Q10 deleted.

Table 16.61 shows an acceptable level of internal consistency across those partner countries and economies that administered questions for the educational career indices with only the scale indicating students' reported activities to obtain career information having a slightly lower reliability.

Table 16.61 Scale reliabilities for educational career indices in partner countries and economies

	INFOCAR	INFOJOB1	INFOJOB2	HOSTCUL	HERITCUL	CULTDIST
Croatia	0.63**	0.75	0.67	*	*	*
Hong Kong-China	0.62	0.67	0.67	0.72	0.81	0.81
Latvia	0.60	0.64	0.60	*	*	*
Macao-China	0.74	0.73	0.74	*	*	*
Serbia	0.74	0.79	0.75	0.88	0.88	0.50
Shanghai-China	0.60***	0.72	0.66	*	*	*
Singapore	0.66***	0.79	0.77	*	*	*

* All questions from the index deleted by the country/economy
** Items EC03Q01 and EC03Q10 deleted by the country.
*** Item EC03Q10 deleted by the country/economy.

Ten items provide data on information about careers (*INFOCAR*). Two response categories were provided "yes", and "no, never". All items were reversed. Table 16.62 shows the item wording and IRT parameters for those countries for which this scale could be created. As can be seen, students reported to use the internet to obtain career information to a much greater extent than they reported visiting a job fair or doing an internship.

Table 16.62 Item parameters for information about careers (*INFOCAR*)

		Parameter estimates
Item	Have you done any of the following to find out about future study or types of work?	Delta
EC03Q01	a) I did an internship	0.81357
EC03Q02	b) I attended <job shadowing or work-site visits>	0.35180
EC03Q03	c) I visited a <job fair>	0.83687
EC03Q04	d) I spoke to a <career advisor> at my school	0.00674
EC03Q05	e) I spoke to a <career advisor> outside of my school	1.23705
EC03Q06	f) I completed a questionnaire to find out about my interests and abilities	-1.10754
EC03Q07	g) I researched the internet for information about careers	-1.47633
EC03Q08	h) I went on an organised tour in an <ISCED 3-5> institution	0.45532
EC03Q09	i) I researched the internet for information about <ISCED 3-5> programmes	-0.72695
EC03Q10	j) <country specific item>	-0.39053

Six items formed the next two indices, which provided data about where students found information about the labour market, either at school (*INFOJOB1*) or outside of school (*INFOJOB2*). The three response categories were "Yes, at school", "Yes, out of school", and "No, never". For the *INFOJOB1* index the initial variables were recoded into "at" variables coded as '1' if the answer was "Yes, at school" and as '0' if the answer was "No, never". For the *INFOJOB2* index the initial variables were recoded into "out" variables coded as '1' if the answer was "Yes, out of school" and as '0' if the answer was "No, never". Tables 16.63 and 16.64 show the item wording and IRT parameters for those countries that administered the items that provided the information for the *INFOJOB1* and *INFOJOB2* respectively.

Table 16.63 Item parameters for information about the labour market provided by the school (*INFOJOB1*)

		Parameter estimates
Item	Which of the following skills have you acquired?	Delta
EC04Q01at	a) How to find information on jobs I am interested in	-1.46392
EC04Q02at	b) How to search for a job	-0.01141
EC04Q03at	c) How to write a <résumé> or a summary of my qualifications	-0.28479
EC04Q04at	d) How to prepare for a job interview	0.92357
EC04Q05at	e) How to find information on <ISCED 3-5> programs I am interested in	-0.45867
EC04Q06at	f) How to find information on student financing (e.g. student loans or grants)	1.29522

Table 16.64 Item parameters for information about the labour market provided outside of school (*INFOJOB2*)

		Parameter estimates
Item	Which of the following skills have you acquired?	Delta
EC04Q01out	a) How to find information on jobs I am interested in	-1.61044
EC04Q02out	b) How to search for a job	-0.69714
EC04Q03out	c) How to write a <résumé> or a summary of my qualifications	0.49007
EC04Q04out	d) How to prepare for a job interview	1.13508
EC04Q05out	e) How to find information on <ISCED 3-5> programs I am interested in	-0.58496
EC04Q06out	f) How to find information on student financing (e.g. student loans or grants)	1.26739

Four items on acculturation in terms of host culture oriented strategies (*HOSTCUL*) were included in the educational career questionnaire. There are four response categories varying from “Strongly agree”, “Agree”, “Disagree” to “Strongly disagree”. All items were reversed. Table 16.65 shows the item wording and the item parameters for those countries that administered the items forming this scale. Results showed that students found it easier to agree with the statement about liking to have a friend from the host culture but harder to agree with the statement about participating in celebrations of the host culture.

Table 16.65 Item parameters for acculturation: Host culture oriented strategies (*HOSTCUL*)

Item	***Below you will find statements about <host culture> and <heritage culture>. <Host culture> refers to the culture and country in which you now live. <Heritage culture> refers to the culture and country in which your mother was born.*** **To what extent do you agree with the following statements?**	Parameter estimates		
		Delta	**tau_1**	**tau_2**
ST23Q01	a) I like to have <host culture> friends	-0.82259	-1.6703	-1.56258
ST23Q03	c) I like to participate in <host culture> celebrations	0.34516	-2.58806	-0.6544
ST23Q05	e) I spend a lot of time with <host culture> friends	-0.14033	-2.50401	-0.5214
ST23Q07	g) I participate in <host culture> celebrations	0.61776	-2.50097	-0.63483

Four items on acculturation in terms of heritage culture oriented strategies (*HERITCUL*) were included in the educational career questionnaire with the four response categories varying from “Strongly agree”, “Agree”, “Disagree” to “Strongly disagree”. All items were reversed. Table 16.66 shows the item wording and the international item parameters for this scale. Again, results showed that it was easier to agree with liking to have friends from the heritage culture than it was to agree with the statement indicating participation in celebrations of the heritage culture.

Table 16.66 Item parameters for acculturation: Heritage culture oriented strategies (*HERITCUL*)

Item	***Below you will find statements about <host culture> and <heritage culture>. <Host culture> refers to the culture and country in which you now live. <Heritage culture> refers to the culture and country in which your mother was born.*** **To what extent do you agree with the following statements?**	Parameter estimates		
		Delta	**tau_1**	**tau_2**
ST23Q02	b) I like to have <heritage culture> friends	-0.73995	-1.90060	-1.19650
ST23Q04	d) I like to participate in <heritage culture> celebrations	-0.09285	-2.15380	-0.69631
ST23Q06	f) I spend a lot of time with <heritage culture> friends	0.46761	-2.17534	-0.39655
ST23Q08	h) I participate in <heritage culture> celebrations	0.36519	-2.06592	-0.56792

Three items on cultural distance between host and heritage culture (*CULTDIST*) were included in the educational career questionnaire. There are four response categories varying from “Strongly agree”, “Agree”, “Disagree” to “Strongly disagree”. All items were reversed. Table 16.67 shows the item wording and the item parameters for those countries which had administered the questions forming this scale.

Table 16.67 Item parameters for cultural distance between host and heritage culture (*CULTDIST*)

Item	***The statements below are about differences between <host culture> and <heritage culture>.*** **To what extent do you agree with the following statements?**	Parameter estimates		
		Delta	**tau_1**	**tau_2**
ST24Q01	a) The values of people in the <host culture> and in the <heritage culture> are the same	-0.37953	-2.20603	-0.04478
ST24Q02	b) Mothers in the <host culture> and in the <heritage culture> treat their children in the same way	0.19734	-2.07529	-0.03418
ST24Q03	c) Pupils from the <host culture> and the <heritage culture> deal with their teachers in the same way	0.18219	-1.87304	-0.13527

SCHOOL QUESTIONNAIRE SCALE INDICES

The School Questionnaire provided data for thirteen scaled indices including four new indices on school leadership (*LEADCOM*, *LEADINST*, *LEADPD*, *LEADTCH*). All indices are listed in Table 16.68.

Table 16.68 School Questionnaire indices

Index	Index label	Relationship to other PISA surveys
School leadership		
LEADCOM	Framing and Communicating the School's Goals and Curricular Development	New
LEADINST	Instructional Leadership	New
LEADPD	Promoting Instructional Improvements and Professional Development	New
LEADTCH	Teacher Participation in Leadership	New
School Autonomy		
SCHAUTON	School autonomy	Used in 2000, 2003 modified for 2012
TCHPARTI	Teacher Participation/Autonomy	Used in 2000, 2003, 2009, modified for 2012
School Resources		
TCSHORT	Shortage of Teaching Staff	Used in 2000, 2003, 2006, 2009
SCMATEDU	Quality of School Educational Resources	Used in 2000, 2003, modified for 2006, 2009
SCMATBUI	Quality of Physical Infrastructure	Used in 2000, 2003
School climate		
STUDCLIM	Student-related Factors Affecting School Climate	New scale; some of the questions were asked in previous cycles
TEACCLIM	Teacher-related Factors Affecting School Climate	New scale; some of the questions were asked in previous cycles
TCMORALE	Teacher Morale	Used in 2000, 2003
TCFOCST	Teacher Focus	New

Tables 16.69 and 16.70 show a high degree of internal consistency for School Questionnaire indices across participating OECD and partner countries and economies respectively.

Table 16.69 Scale reliabilities for School Questionnaire indices in OECD countries

	LEADCOM	LEADINST	LEADPD	LEADTCH	SCHAUTON	TCHPARTI	TCSHORT	SCMATEDU	SCMATBUI	STUDCLIM	TEACCLIM	TCMORALE	TCFOCST
Australia	0.77	0.74	0.81	0.78	0.78	0.72	0.87	0.84	0.82	0.89	0.87	0.81	0.61
Austria	0.69	0.80	0.89	0.73	0.82	0.77	0.78	0.83	0.80	0.85	0.82	0.76	0.48
Belgium	0.68	0.74	0.85	0.75	0.83	0.65	0.82	0.82	0.82	0.89	0.82	0.73	0.58
Canada	0.77	0.78	0.79	0.74	0.70	0.67	0.82	0.81	0.78	0.85	0.83	0.80	0.62
Chile	0.75	0.76	0.81	0.72	0.88	0.72	0.89	0.85	0.81	0.90	0.89	0.81	0.67
Czech Republic	0.76	0.82	0.84	0.78	0.96	0.74	0.74	0.78	0.75	0.86	0.81	0.73	0.64
Denmark	0.78	0.67	0.79	0.73	0.81	0.73	0.71	0.76	0.79	0.86	0.86	0.85	0.48
Estonia	0.74	0.68	0.76	0.64	0.76	0.71	0.56	0.71	0.80	0.82	0.82	0.79	0.68
Finland	0.77	0.74	0.81	0.73	0.76	0.69	0.75	0.79	0.86	0.83	0.82	0.73	0.63
France	0.73	0.64	0.78	0.79	0.77	0.53	0.81	0.82	0.82	0.86	0.79	0.74	0.70
Germany	0.68	0.72	0.77	0.63	0.69	0.70	0.77	0.81	0.81	0.81	0.76	0.79	0.33
Greece	0.78	0.79	0.78	0.79	0.52	0.42	0.89	0.76	0.87	0.88	0.87	0.78	0.52
Hungary	0.69	0.72	0.75	0.66	0.78	0.68	0.69	0.76	0.79	0.88	0.74	0.78	0.62
Iceland	0.70	0.72	0.82	0.70	0.73	0.69	0.67	0.77	0.76	0.86	0.82	0.77	0.55
Ireland	0.74	0.72	0.77	0.78	0.81	0.67	0.60	0.75	0.82	0.84	0.86	0.83	0.59
Israel	0.67	0.72	0.76	0.75	0.77	0.74	0.83	0.87	0.79	0.89	0.90	0.79	0.54
Italy	0.79	0.75	0.80	0.80	0.68	0.57	0.85	0.81	0.81	0.83	0.85	0.72	0.60
Japan	0.68	0.71	0.76	0.69	0.58	0.83	0.85	0.82	0.69	0.89	0.85	0.78	0.46
Korea	0.83	0.82	0.83	0.84	0.79	0.79	0.94	0.88	0.80	0.90	0.90	0.82	0.70
Luxembourg	0.39	0.70	0.82	0.74	0.73	0.64	0.81	0.73	0.81	0.86	0.83	0.67	0.68
Mexico	0.82	0.77	0.79	0.82	0.78	0.57	0.87	0.88	0.81	0.84	0.88	0.86	0.70
Netherlands	0.75	0.69	0.79	0.73	0.54	0.66	0.70	0.81	0.76	0.84	0.75	0.77	0.61
New Zealand	0.79	0.73	0.80	0.79	0.66	0.72	0.84	0.82	0.84	0.90	0.87	0.83	0.59
Norway	0.80	0.77	0.79	0.69	0.63	0.65	0.78	0.75	0.83	0.80	0.82	0.82	0.55
Poland	0.72	0.65	0.83	0.71	0.62	0.47	0.33	0.78	0.68	0.82	0.82	0.78	0.51
Portugal	0.70	0.79	0.75	0.80	0.80	0.60	0.84	0.78	0.79	0.87	0.85	0.76	0.50
Slovak Republic	0.70	0.68	0.73	0.76	0.61	0.69	0.61	0.75	0.85	0.81	0.72	0.72	0.29
Slovenia	0.76	0.85	0.81	0.79	0.77	0.65	0.74	0.81	0.84	0.86	0.79	0.77	0.66
Spain	0.69	0.74	0.79	0.77	0.84	0.61	0.78	0.78	0.81	0.85	0.86	0.72	0.69
Sweden	0.79	0.73	0.80	0.68	0.68	0.61	0.73	0.69	0.89	0.85	0.86	0.75	0.60
Switzerland	0.63	0.65	0.80	0.71	0.74	0.64	0.76	0.77	0.79	0.81	0.81	0.74	0.54
Turkey	0.77	0.81	0.84	0.88	0.74	0.74	0.88	0.83	0.75	0.87	0.89	0.78	0.77
United Kingdom	0.79	0.75	0.80	0.77	0.67	0.62	0.81	0.87	0.84	0.85	0.87	0.84	0.47
United States	0.80	0.80	0.83	0.80	0.76	0.71	0.85	0.86	0.74	0.87	0.90	0.83	0.68
OECD median	**0.75**	**0.74**	**0.80**	**0.75**	**0.76**	**0.68**	**0.79**	**0.81**	**0.81**	**0.86**	**0.84**	**0.78**	**0.60**

Table 16.70 Scale reliabilities for school questionnaire indices in partner countries and economies

	LEADCOM	LEADINST	LEADPD	LEADTCH	SCHAUTON	TCHPARTI	TCSHORT	SCMATEDU	SCMATBUI	STUDCLIM	TEACCLIM	TCMORALE	TCFOCST
Albania	0.56	0.55	0.76	0.73	0.63	0.55	0.80	0.76	0.70	0.79	0.82	0.74	0.70
Argentina	0.76	0.81	0.78	0.74	0.72*	0.53*	0.87	0.81	0.84	0.85	0.86	0.74	0.61
Brazil	0.78	0.72	0.75	0.80	0.81	0.67	0.84	0.82	0.82	0.85	0.87	0.85	0.70
Bulgaria	0.62	0.65	0.61	0.75	0.64	0.67	0.20	0.77	0.71	0.88	0.89	0.69	0.49
Colombia	0.84	0.73	0.82	0.78	0.87	0.71	0.92	0.91	0.91	0.90	0.90	0.84	0.59
Costa Rica	0.84	0.78	0.83	0.84	0.78	0.53	0.79	0.87	0.85	0.87	0.85	0.83	0.68
Croatia	0.76	0.77	0.75	0.76	0.74	0.68	0.65	0.73	0.75	0.86	0.79	0.79	0.69
Cyprus[1, 2]	0.70	0.71	0.76	0.74	0.88	0.65	0.96	0.86	0.86	0.88	0.88	0.84	0.46
Hong Kong-China	0.76	0.77	0.77	0.85	0.71	0.73	0.87	0.86	0.80	0.87	0.88	0.79	0.52
Indonesia	0.81	0.73	0.73	0.79	0.81	0.81	0.85	0.88	0.64	0.74	0.80	0.84	0.68
Jordan	0.78	0.67	0.80	0.81	0.82	0.75	0.92	0.82	0.84	0.88	0.90	0.85	0.75
Kazakhstan	0.64	0.61	0.65	0.76	0.76	0.66	0.91	0.85	0.84	0.93	0.94	0.79	0.52
Latvia	0.73	0.67	0.81	0.73	0.74	0.69	0.76	0.72	0.73	0.85	0.83	0.73	0.58
Liechtenstein	0.88	0.85	0.89	0.73	0.84	0.81	0.74	0.32	0.81	0.62	0.79	0.76	0.25
Lithuania	0.59	0.72	0.84	0.73	0.77	0.77	0.57	0.73	0.75	0.81	0.78	0.76	0.68
Macao-China	0.64	0.68	0.70	0.56	0.93	0.82	0.92	0.89	0.84	0.94	0.94	0.84	0.41
Malaysia	0.85	0.79	0.80	0.79	0.77	0.81	0.70	0.84	0.83	0.88	0.82	0.88	0.64
Montenegro	0.84	0.81	0.84	0.80	0.70	0.50	0.54	0.73	0.66	0.83	0.79	0.80	0.65
Peru	0.82	0.79	0.72	0.76	0.87	0.67	0.87	0.89	0.82	0.76	0.89	0.83	0.71
Qatar	0.78	0.75	0.79	0.84	0.89	0.94	0.89	0.85	0.83	0.87	0.88	0.81	0.69
Romania	0.86	0.91	0.84	0.85	0.65	0.49	0.66	0.76	0.56	0.77	0.79	0.66	0.42
Russian Federation	0.66	0.72	0.79	0.71	0.81	0.68	0.84	0.83	0.78	0.91	0.90	0.74	0.51
Serbia	0.74	0.75	0.82	0.72	0.70	0.58	0.66	0.82	0.79	0.85	0.84	0.70	0.81
Shanghai-China	0.58	0.73	0.68	0.71	0.90	0.82	0.92	0.91	0.83	0.94	0.94	0.80	0.63
Singapore	0.83	0.78	0.84	0.80	0.79	0.77	0.76	0.86	0.79	0.88	0.87	0.86	0.57
Chinese Taipei	0.84	0.78	0.85	0.86	0.82	0.77	0.92	0.92	0.86	0.92	0.93	0.88	0.64
Thailand	0.83	0.83	0.84	0.85	0.86	0.86	0.86	0.90	0.87	0.84	0.82	0.85	0.73
Tunisia	0.63	0.73	0.71	0.81	0.79	0.62	0.84	0.80	0.77	0.83	0.79	0.83	0.67
United Arab Emirates	0.79	0.76	0.79	0.79	0.92	0.84	0.95	0.89	0.89	0.91	0.92	0.88	0.73
Uruguay	0.76	0.65	0.70	0.83	0.83	0.66	0.84	0.77	0.85	0.89	0.89	0.74	0.75
Viet Nam	0.65	0.79	0.73	0.74	0.69	0.67	0.92	0.86	0.78	0.80	0.77	0.81	0.62
Median	**0.76**	**0.75**	**0.79**	**0.78**	**0.79**	**0.68**	**0.84**	**0.84**	**0.82**	**0.87**	**0.87**	**0.81**	**0.64**

* SC33Q01 was deleted by the country.

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

School leadership

In 2012, the PISA School Questionnaire contained 21 items about school leadership 13 of which provided data for four scaled indices. Principals were asked to indicate the frequency of the listed activities and behaviours in their school during the last academic year. The six response categories were "Did not occur", "1-2 times during the year", "3-4 times during the year", "Once a month", "Once a week", to "More than once a week".

Table 16.71 shows the item wording and the international item parameters for framing and communicating the school's goals and curricular development (*LEADCOM*).

Table 16.71 Item parameters for framing and communicating the school's goals and curricular development (*LEADCOM*)

Item	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during <the last academic year>.	Parameter estimates				
		Delta	tau_1	tau_2	tau_3	tau_4
SC34Q02	b) I use student performance results to develop the school's educational goals	0.31712	-2.73434	-0.48319	0.82003	1.64976
SC34Q03	c) I make sure that the professional development activities of teachers are in accordance with the teaching goals of the school	0.11488	-2.76262	-0.43718	0.54590	1.62237
SC34Q04	d) I ensure that teachers work according to the school's educational goals	-0.51262	-2.75791	-0.45630	0.55446	1.57012
SC34Q14	n) I discuss the school's academic goals with teachers at faculty meetings	0.08100	-3.10596	-1.10527	0.23969	1.96756

Table 16.72 shows the item wording and the international item parameters for instructional leadership (*LEADINST*).

Table 16.72 Item parameters for instructional leadership (*LEADINST*)

Item	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during <the last academic year>.	Parameter estimates				
		Delta	tau_1	tau_2	tau_3	tau_4
SC34Q05	e) I promote teaching practices based on recent educational research	0.38344	-2.46672	-0.40380	0.40411	1.57547
SC34Q06	f) I praise teachers whose students are actively participating in learning	-0.27321	-2.21641	-0.76191	0.25683	1.38666
SC34Q08	h) I draw teachers' attention to the importance of pupil's development of critical and social capacities	-0.11000	-2.51090	-0.71637	0.31297	1.45357

Table 16.73 shows the item wording and the international item parameters for promoting instructional improvements and professional development (*LEADPD*).

Table 16.73 Item parameters for promoting instructional improvements and professional development (*LEADPD*)

Item	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during <the last academic year>.	Parameter estimates				
		Delta	tau_1	tau_2	tau_3	tau_4
SC34Q07	g) When a teacher has problems in his/her classroom, I take the initiative to discuss matters	0.27416	-2.71542	-0.96912	0.15629	1.52112
SC34Q09	i) I pay attention to disruptive behaviour in classrooms	-0.22389	-2.01263	-0.83660	-0.01204	1.19856
SC34Q13	m) When a teacher brings up a classroom problem, we solve the problem together	-0.05000	-2.63794	-1.08801	0.15253	1.50377

Table 16.74 shows the item wording and the international item parameters for teacher participation in leadership (*LEADTCH*). It can be seen that one item is notably more difficult than other items in the scale. This result indicates that principals reported asking the teachers in their schools to participate in reviewing management practices less frequently than in other activities.

Table 16.74 Item parameters for teacher participation in leadership (*LEADTCH*)

Item	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during <the last academic year>.	Parameter estimates				
		Delta	tau_1	tau_2	tau_3	tau_4
SC34Q10	j) I provide staff with opportunities to participate in school decision-making	-0.38465	-2.80658	-1.23150	0.27523	1.79014
SC34Q11	k) I engage teachers to help build a school culture of continuous improvement	-0.53890	-2.76802	-0.97913	0.41103	1.63530
SC34Q12	l) I ask teachers to participate in reviewing management practices	0.92400	-2.63863	-0.35591	0.18107	1.50275

School autonomy

In 2012, the PISA School Questionnaire contained twelve items about school autonomy. Principals were asked to indicate who had a considerable responsibility for various tasks in their school. There were five response categories "Principal", "Teachers", "<School governing board>“, "<Regional or local education authority>“, "National education authority" and principals were asked to tick as many categories as appropriate. All twelve items provided data for two scaled indices, namely school autonomy (*SCHAUTON*) and teacher participation (*TCHPARTI*). However, these data were recoded differently for each index.

Table 16.75 shows the item wording and the international item parameters for school autonomy (*SCHAUTON*). The items were recoded as follows: If at least one of the categories "Principal", "Teachers" or "School governing board“ was ticked, the scaled variable was coded as '1', otherwise it was coded as '0'. The easiest item was "Establishing student disciplinary policies" (δ =-2.06), indicating that this was usually done within the school. The most difficult item was "Establishing teachers' starting salaries" (δ =2.88) indicating that this was usually done by local or national authorities.

Table 16.75 Item parameters for school autonomy (*SCHAUTON*)

Item	Regarding your school, who has a considerable responsibility for the following tasks?	Parameter estimates
		Delta
SC33Q01	a) Selecting teachers for hire	-0.08692
SC33Q02	b) Firing teachers	0.47358
SC33Q03	c) Establishing teachers' starting salaries	2.87671
SC33Q04	d) Determining teachers' salary increases	2.49770
SC33Q05	e) Formulating the school budget	0.15668
SC33Q06	f) Deciding on budget allocations within the school	-1.59541
SC33Q07	g) Establishing student disciplinary policies	-2.05562
SC33Q08	h) Establishing student assessment policies	-0.85521
SC33Q09	i) Approving students for admission to the school	-0.77633
SC33Q10	j) Choosing which textbooks are used	-0.91927
SC33Q11	k) Determining course content	0.28079
SC33Q12	l) Deciding which courses are offered	0.00300

Table 16.76 shows the item wording and the international parameters used for teacher participation (*TCHPARTI*) in school decisions. The recoding was based on the "Teachers" column only. A "tick" in this column was coded as '1' and no "tick" was coded as '0'. The distribution of item difficulties for this scale was reasonable and appropriate. The decisions relating to teachers themselves (e.g. firing teachers or hiring teachers and establishing their salaries) were difficult items, indicating that usually teachers do not participate in these decisions as could be expected. In contrast, the item regarding the choice of textbooks was a very easy item, indicating that usually teachers participated in this decision.

Table 16.76 Item parameters for teacher participation (*TCHPARTI*)

Item	Regarding your school, who has a considerable responsibility for the following tasks?	Parameter estimates Delta
SC33Q01	a) Selecting teachers for hire	1.24783
SC33Q02	b) Firing teachers	2.47040
SC33Q03	c) Establishing teachers' starting salaries	3.40994
SC33Q04	d) Determining teachers' salary increases	2.79464
SC33Q05	e) Formulating the school budget	1.45142
SC33Q06	f) Deciding on budget allocations within the school	0.67972
SC33Q07	g) Establishing student disciplinary policies	-2.29624
SC33Q08	h) Establishing student assessment policies	-2.75098
SC33Q09	i) Approving students for admission to the school	0.64240
SC33Q10	j) Choosing which textbooks are used	-3.62473
SC33Q11	k) Determining course content	-2.61041
SC33Q12	l) Deciding which courses are offered	-1.41400

School resources

The PISA 2012 school questionnaire contained thirteen items about school resources, measuring the school principal's perceptions of potential factors hindering instruction at school. The four response categories were "Not at all", "Very little", "To some extent", to "A lot".

The index on teacher shortage (*TCSHORT*) was derived from four items. Similar items were used in PISA 2000, 2003, 2006 and 2009. The items were not reversed for scaling. Table 16.77 shows the item wording and the international parameters used for IRT scaling. As was the case in 2009, the deltas indicated that principals found it harder to recruit teachers who were appropriately qualified in the test language than in mathematics.

Table 16.77 Item parameters for teacher shortage (*TCSHORT*)

		Parameter estimates		
Item	Is your school's capacity to provide instruction hindered by any of the following issues?	Delta	tau_1	tau_2
SC14Q01	a) A lack of qualified science teachers	0.02345	-1.51534	-0.51175
SC14Q02	b) A lack of qualified mathematics teachers	-0.00025	-1.39991	-0.33669
SC14Q03	c) A lack of qualified <test language> teachers	0.50620	-1.59703	-0.25525
SC14Q04	d) A lack of qualified teachers of other subjects	-0.52900	-2.59607	-0.28596

The index on the school's educational resources (*SCMATEDU*) was computed on the basis of six items measuring the school principals' perceptions of potential factors hindering instruction at school. Similar items were used in PISA 2000 and 2003 but question format and item wording were modified for PISA 2006 and PISA 2009. For 2012 the items were modified from 2009. All items were reversed for scaling. Table 16.78 shows the item wording and the international parameters used for IRT scaling.

Table 16.78 Item parameters for quality of educational resources (*SCMATEDU*)

		Parameter estimates		
Item	Is your school's capacity to provide instruction hindered by any of the following issues?	Delta	tau_1	tau_2
SC14Q05	e) Shortage or inadequacy of science laboratory equipment	0.33310	-1.31836	0.17403
SC14Q06	f) Shortage or inadequacy of instructional materials (e.g. textbooks)	-0.49770	-1.70343	0.19337
SC14Q07	g) Shortage or inadequacy of computers for instruction	0.18530	-1.46462	0.25553
SC14Q08	h) Lack or inadequacy of Internet connectivity	-0.20079	-1.06959	0.15556
SC14Q09	i) Shortage or inadequacy of computer software for instruction	0.20494	-1.79123	0.18129
SC14Q10	j) Shortage or inadequacy of library materials	0.00015	-1.70562	0.07394

The index concerning the quality of physical infrastructure (*SCMATBUI*) was computed on the basis of three items measuring the principals' perceptions of potential factors hindering instruction at school. Similar items were used in PISA 2000 and 2003 but question format and item wording were modified for PISA 2006 and PISA 2009. For 2012 the items were modified from 2009. All items were reversed for scaling. Table 16.79 shows the item wording and the international parameters used for IRT scaling.

Table 16.79 Item parameters for quality of physical infrastructure (*SCMATBUI*)

		Parameter estimates		
Item	Is your school's capacity to provide instruction hindered by any of the following issues?	Delta	tau_1	tau_2
SC14Q11	k) Shortage or inadequacy of school buildings and grounds	0.29073	-1.53165	0.29387
SC14Q12	l) Shortage or inadequacy of heating/cooling and lighting systems	-0.25216	-1.33150	0.09259
SC14Q13	m) Shortage or inadequacy of instructional space (e.g. classrooms)	-0.03900	-1.58587	0.21897

School climate

In 2012, the PISA School Questionnaire contained three items batteries about school climate. The first item battery containing nineteen items measured the school principals' perceptions of potential phenomena hindering instruction at school. The four response categories were "Not at all", "Very little", "To some extent" and "A lot". This item battery contributed to two indices, namely student-related factors affecting school climate (*STUDCLIM*) and teacher related factors affecting school climate (*TEACCLIM*). Two other item batteries measured school principals' endorsement of statements regarding teacher morale (*TCMORALE*) and teacher focus on students (*TCFOCST*). The four response categories for endorsement ranged from "Strongly agree", "Agree", "Disagree" to "Strongly disagree".

Eight items regarding student related aspects of school climate, which had been employed in previous PISA cycles, were used for the index regarding the student-related aspects of school climate. Table 16.80 shows the item wording and the international parameters used for IRT scaling.

Table 16.80 Item parameters for student-related aspects of school climate (*STUDCLIM*)

		Parameter estimates		
Item	In your school, to what extent is the learning of students hindered by the following phenomena?	Delta	tau_1	tau_2
SC22Q01	a) Student truancy	-0.76575	-2.30414	0.25299
SC22Q02	b) Students skipping classes	-0.57107	-2.50950	0.29456
SC22Q03	c) Students arriving late for school	-0.59703	-3.15841	0.37340
SC22Q04	d) Students not attending compulsory school events (e.g. sports day) or excursions	0.66621	-2.87294	0.41253
SC22Q05	e) Students lacking respect for teachers	0.10665	-2.66405	0.44452
SC22Q06	f) Disruption of classes by students	-0.42016	-2.82902	0.27133
SC22Q07	g) Student use of alcohol or illegal drugs	1.02770	-1.61249	0.92065
SC22Q08	h) Students intimidating or bullying other students	0.55300	-2.71971	0.86493

Eleven questions regarding teacher-related factors affecting school climate has appeared in previous PISA cycles and were used for the index on the teacher-related factors affecting school climate in PISA 2012. Table 16.81 shows the item wording and the international parameters used for IRT scaling.

Table 16.81 Item parameters for teacher-related factors affecting school climate (*TEACCLIM*)

		Parameter estimates		
Item	In your school, to what extent is the learning of students hindered by the following phenomena?	Delta	tau_1	tau_2
SC22Q09	i) Students not being encouraged to achieve their full potential	-0.15258	-1.68163	-0.09457
SC22Q10	j) Poor student-teacher relations	0.38347	-2.13781	0.86510
SC22Q11	k) Teachers having to teach students of heterogeneous ability levels within the same class	-1.39607	-1.75757	-0.29288
SC22Q12	l) Teachers having to teach students of diverse ethnic backgrounds (i.e. language, culture) within the same class	0.11387	-1.09272	0.00761
SC22Q13	m) Teachers' low expectations of students	0.26123	-1.91961	-0.04530
SC22Q14	n) Teachers not meeting individual students' needs	-0.09722	-2.23969	0.05272
SC22Q15	o) Teacher absenteeism	0.00719	-1.73082	0.51369
SC22Q16	p) Staff resisting change	-0.15440	-1.95268	0.05339
SC22Q17	q) Teachers being too strict with students	0.50292	-2.55852	0.33750
SC22Q18	r) Teachers being late for classes	0.40059	-1.97713	0.76795
SC22Q19	s) Teachers not being well prepared for classes	0.13100	-1.91099	0.85873

Four items on teacher morale (*TCMORALE*) were included in the School Questionnaire. There are four response categories varying from "Strongly agree", "Agree", "Disagree" to "Strongly disagree". All items were reversed. Table 16.82 shows the item wording and the international item parameters for this scale.

Table 16.82 Item parameters for teacher morale (*TCMORALE*)

		Parameter estimates		
Item	**Think about the teachers in your school. How much do you agree with the following statements?**	**Delta**	**tau_1**	**tau_2**
SC26Q01	a) The morale of teachers in this school is high	0.34116	-3.83956	-0.79275
SC26Q02	b) Teachers work with enthusiasm	0.40178	-4.13394	-1.00884
SC26Q03	c) Teachers take pride in this school	-0.15802	-3.86018	-0.82074
SC26Q04	d) Teachers value academic achievement	-0.58500	-3.57413	-0.97397

Three items measuring teacher focus (*TCFOCST*) on students were included in the School Questionnaire which were based on item batteries regarding teacher consensus in earlier cycles that had been shown to measure more the disagreement among teachers regarding certain pedagogical issues rather than consensus. The four response categories ranged from "Strongly agree" to "Strongly disagree". All items were reversed. Table 16.83 shows the item wording and the international item parameters for this scale.

Table 16.83 Item parameters for teacher focus (*TCFOCST*)

		Parameter estimates		
Item	**How much do you agree with these statements about teachers in your school?**	**Delta**	**tau_1**	**tau_2**
SC27Q01	a) Mathematics teachers are interested in trying new methods and teaching practices	-0.32068	-3.16187	-0.26171
SC28Q02	d) There is consensus among mathematics teachers that it is best to adapt academic standards to the students' levels and needs	0.26357	-2.62849	-0.20707
SC29Q01	e) There is consensus among mathematics teachers that the social and emotional development of the students is as important as their acquisition of mathematical skills and knowledge in mathematics classes	0.05700	-3.03557	-0.09006

Parent Questionnaire scale indices

Parent Questionnaire indices were only available for those 11 countries that chose to administer the optional parent questionnaire. The data from the Parent Questionnaire contributed to the five indices that are listed in Table 16.84. All indices except parents' perception of school quality (*PQSCHOOL*) were modified from the previous PISA surveys, only *PQSCHOOL* is used the same way as in 2006 and 2009.

Table 16.84 Parent Questionnaire indices

Index	Index label	Relationship to other PISA surveys
PQSCHOOL	Parents' perception of school quality	Used in 2006 and 2009
PARINVOL	Parental involvement in their child's school	More items were added to 2009 scale
PARSUPP	Student support	Modified from 2009
PQMIMP	Parent attitudes toward mathematics	Modified from 2006 to be mathematics specific
PQMCAR	Mathematics career	Modified from 2006 to be mathematics specific

Tables 16.85 and 16.86 show very acceptable degree of internal consistency across participating OECD and partner countries and economies for the Parent Questionnaire indices.

Table 16.85 Scale reliabilities for Parent Questionnaire indices in OECD countries

	PQSCHOOL	PARINVOL	PARSUPP	PQMIMP	PQMCAR
Belgium	0.81	0.55	0.70	0.85	0.81
Chile	0.86	0.67	0.77	0.86	0.81
Germany	0.83	0.64	0.73	0.87	0.60
Hungary	0.85	0.68	0.73	0.85	0.79
Italy	0.83	0.63*	0.65	0.85	0.79
Korea	0.87	0.71	0.78	0.84	0.78
Mexico	0.86	0.75	0.79	0.86	0.74
Portugal	0.85	0.68	0.73	0.87	0.79

* PA10Q06 and PA10Q11 deleted by the country.

Table 16.86 Scale reliabilities for Parent Questionnaire indices in partner countries and economies

	PQSCHOOL	PARINVOL	PARSUPP	PQMIMP	PQMCAR
Croatia	0.81	0.58*	0.68	0.83	0.69
Hong Kong-China	0.83	0.73	0.80	0.86	0.72
Macao-China	0.85	0.75	0.81	0.86	0.76

* PA10Q06 and PA10Q11 deleted by the country.

Seven items measuring parents' perceptions of the quality of school learning were included in the PISA 2012 Parent Questionnaire as was the case in PISA 2006 and PISA 2009. Parents were asked how much they agreed with the seven statements. The response categories included "Strongly agree", "Agree", "Disagree" and "Strongly disagree". The items were reverse coded for scaling. The item wording and international parameters for IRT scaling are shown in Table 16.87. Results for the item deltas indicated that it was harder for parents to agree that the school their child attended had high standards of achievement and easier to agree that teachers seemed competent and dedicated.

Table 16.87 Item parameters for parent's perception of school quality (*PQSCHOOL*)

Item	How much do you agree or disagree with the following statements?	Parameter estimates Delta	tau_1	tau_2
PA09Q01	a) Most of my child's school teachers seem competent and dedicated	-0.46021	-3.12016	-1.15967
PA09Q02	b) Standards of achievement are high in my child's school	0.46438	-3.65406	-0.26069
PA09Q03	c) I am happy with the content taught and the instructional methods used in my child's school	-0.06860	-3.29511	-0.96367
PA09Q04	d) I am satisfied with the disciplinary atmosphere in my child's school	0.04788	-2.59975	-0.89333
PA09Q05	e) My child's progress is carefully monitored by the school	0.07441	-3.35336	-0.62689
PA09Q06	f) My child's school provides regular and useful information on my child's progress	0.32348	-2.84613	-0.63228
PA09Q07	g) My child's school does a good job in educating students	-0.38134	-2.98083	-1.00424

The scale regarding parental involvement was modified from PISA 2009 in that more items were added to align with a similar item battery in the school questionnaire. In 2012, eleven items measured parents' involvement in their child's school. The parents were asked whether they had participated in various school-related activities during the previous academic year. The response categories were "Yes" and "No". The items were reverse coded for scaling. The item wording and international parameters for IRT scaling are shown in Table 16.88. The distribution of item difficulties for this scale was reasonable and appropriate with easy items such as discussing the child's behaviour or progress and more difficult ones such as appearing as a guest speaker or volunteering in the school canteen.

Table 16.88 Item parameters for parental involvement (*PARINVOL*)

Item	During the last <academic year>, have you participated in any of the following school-related activities?	Parameter estimates Delta
PA10Q01	a) Discussed my child's behaviour with a teacher on my own initiative	-2.23093
PA10Q02	b) Discussed my child's behaviour on the initiative of one of his/her teachers	-1.88603
PA10Q03	c) Volunteered in physical activities, e.g. building maintenance, carpentry, gardening or yard work	1.00116
PA10Q04	d) Volunteered in extra-curricular activities, e.g. book club, school play, sports, field trip	0.29876
PA10Q05	e) Volunteered in the school library or media centre	1.63260
PA10Q06	f) Assisted a teacher in the school	0.86205
PA10Q07	g) Appeared as a guest speaker	1.96006
PA10Q08	h) Participated in local school <government>, e.g. parent council or school management committee	0.04862
PA10Q09	i) Discussed my child's progress with a teacher on my own initiative	-2.28397
PA10Q10	j) Discussed my child's progress on the initiative of one of their teachers	-1.88242
PA10Q11	k) Volunteered in the school <canteen>	2.48010

Seven items measuring support of students by their parents were included in the PISA 2012 Parent Questionnaire. Some items were modified from PISA 2009 to be mathematics specific. Parents were asked how often they or someone else in their home did various things with their child. The response categories included "Never or hardly ever", "Once or twice a year", "Once or twice a month", "Once or twice a week" and "Every day or almost every day". The item wording and international parameters for IRT scaling are shown in Table 16.89. Item deltas indicated that parents more frequently had the main meal with their child and less frequently helped their child with his or her mathematics homework.

Table 16.89 Item parameters for student support (*PARSUPP*)

Item	How often do you or someone else in your home do the following things with your child?	Parameter estimates			
		Delta	tau_1	tau_2	tau_3
PA13Q01	a) Discuss how well my child is doing at school	-0.87051	-0.51519	-0.47941	0.15863
PA13Q02	b) Eat <the main meal> with my child around a table	-1.41782	0.70684	-0.13095	-0.88945
PA13Q03	c) Spend time just talking to my child	-1.13481	0.32403	-0.75715	-0.38028
PA13Q04	d) Help my child with his/her mathematics homework	1.25815	0.43140	-1.21920	-0.31943
PA13Q05	e) Discuss how my child is performing in mathematics class	0.16986	0.04751	-1.14690	-0.12075
PA13Q06	f) Obtain mathematics materials (e.g. applications, software, study guides etc.) for my child	1.24696	-0.36108	-0.39858	0.12684
PA13Q07	g) Discuss with my child how mathematics can be applied in everyday life	0.74817	-0.23617	-0.76437	0.25283

Four items measuring parent attitudes toward mathematics were included in the PISA 2012 Parent Questionnaire. The items were modified from PISA 2006 to be mathematics specific. Parents were asked how much they agreed with the four statements. The response categories ranged from "Strongly agree" to "Strongly disagree". The items were reverse coded for scaling. The item wording and international parameters for IRT scaling are shown in Table 16.90.

Table 16.90 Item parameters for parent attitudes toward mathematics (*PQMIMP*)

Item	How much do you agree with the following statements?	Parameter estimates		
		Delta	tau_1	tau_2
PA14Q01	a) It is important to have good mathematics knowledge and skills in order to get any good job in today's world	-0.37976	-3.21500	-0.66593
PA14Q02	b) Employers generally appreciate strong mathematics knowledge and skills among their employees	0.60015	-4.00749	-0.27891
PA14Q03	c) Most jobs today require some mathematics knowledge and skills	0.10841	-3.98955	-0.58688
PA14Q04	d) It is an advantage in the job market to have good mathematics knowledge and skills	-0.32880	-3.26157	-0.78352

Five items measuring parental expectations regarding their child's and their own involvement in mathematics career were included in the PISA 2012 parent questionnaire. The items were modified from PISA 2006 to be mathematics specific. The response categories were "Yes" and "No" and all items were reverse coded for scaling. The item wording and international parameters for IRT scaling are shown in Table 16.91. It was of interest to note that the item deltas indicated that parents more easily affirmed that they expected their children to go into a mathematics related career yet they had not seen their child showing interest in studying mathematics after leaving secondary school.

Table 16.91 Item parameters for mathematics career (*PQMCAR*)

Item	Please answer the questions below	Parameter estimates
		Delta
PA15Q01	a) Does anybody in your family (including you) work in a <mathematics-related career>?	-0.02677
PA15Q02	b) Does your child show an interest in working in a <mathematics-related career>?	-0.09525
PA15Q03	c) Do you expect your child will go into a <mathematics-related career>?	-0.52103
PA15Q04	d) Has your child shown interest in studying mathematics after completing <secondary school>?	0.72876
PA15Q05	e) Do you expect your child will study mathematics after completing <secondary school>?	-0.08571

The *PISA index of economic, social and cultural status* (ESCS)

Computation of ESCS

The index of ESCS was used first in the PISA 2000 analysis and at that time was derived from five indices: highest occupational status of parents (*HISEI*), highest educational level of parents (in years of education according to ISCED), family wealth, cultural possessions and home educational resources (all three WLEs based on student reports on home possessions).

The ESCS for PISA 2003 and 2006 was derived from three variables related to family background: highest parental education (in number of years of education according to ISCED classification), highest parental occupation (*HISEI* scores), and number of home possessions including books in the home. The rationale for using these three components was that socio-economic status has usually been seen as based on education, occupational status and income. As no direct income measure has been available from the PISA data, the existence of household items has been used as a proxy for family wealth.

The ESCS was slightly modified in PISA 2009 because: *(i)* more indicators were available in that survey; and (ii) a consultation with countries regarding the mapping of ISCED levels to years of schooling led to minor changes in the indicator of parental education.

As in PISA 2003, PISA 2006, and PISA 2009, the variables comprising ESCS for PISA 2012 included: home possessions (*HOMEPOS*) – which comprised all items on the *WEALTH*, *CULTPOS* and *HEDRES* scales, as well as books in the home (ST28Q01) recoded into a four-level categorical variable (fewer than or equal to 25 books, 26-100 books, 101-500 books,

more than 500 books); the highest parental occupation (*HISEI*); and the highest parental education expressed as years of schooling (*PARED*). However, the home possessions scale for PISA 2012 was computed differently than in the previous cycles for the purpose of enabling a trend study. For more details please refer to the section on trends in ESCS below.

Missing values for students with missing data for only one variable were imputed with predicted values plus a random component based on a regression on the other two variables. If there were missing data on more than one variable, ESCS was not computed for that case and a missing value was assigned for ESCS. Variables with imputed values were then used for a principal component analysis with an OECD senate weight.

The ESCS scores were obtained as component scores for the first principal component with zero being the score of an average OECD student and one being the standard deviation across equally weighted OECD countries. For partner countries and economies, ESCS scores were obtained as

16.5

$$ESCS = \frac{\beta_1 HISEI' + \beta_2 PARED' + \beta_3 HOMEPOS'}{\varepsilon_f}$$

where β_1, β_2 and β_3 are the OECD factor loadings, *HISEI'*, *PARED'* and *HOMEPOS'* the "OECD-standardised" variables and ε_f is the eigenvalue of the first principal component.[3]

Consistency across countries

Using principal component analysis (PCA) to derive factor loading for each participating country provided insight into the extent to which relationships were similar between the three variables. Table 16.92 shows the PCA results for the OECD countries and Table 16.93 shows those for partner countries and economies. The tables also include the scale reliabilities (Cronbach's alpha) for the *z*-standardised variables.

Comparing results from within-country PCA revealed generally similar patterns of factor loadings across countries. Only in a few countries somewhat distinct patterns emerged, however, all three variables contributed more or less equally to this index. The median scale reliability for the pooled OECD countries was 0.65.

Table 16.92 Factor loadings and internal consistency of ESCS 2012 in OECD countries

	Factor loadings			
	HISEI	PARED	HOMEPOS	Reliability
Australia	0.78	0.78	0.67	0.57
Austria	0.80	0.81	0.67	0.65
Belgium	0.84	0.81	0.70	0.69
Canada	0.79	0.79	0.66	0.60
Chile	0.89	0.88	0.84	0.79
Czech Republic	0.81	0.81	0.65	0.59
Denmark	0.83	0.80	0.72	0.67
Estonia	0.83	0.79	0.67	0.65
Finland	0.80	0.79	0.67	0.60
France	0.80	0.78	0.70	0.62
Germany	0.84	0.81	0.64	0.65
Greece	0.85	0.84	0.70	0.72
Hungary	0.84	0.85	0.74	0.75
Iceland	0.81	0.80	0.59	0.57
Ireland	0.81	0.81	0.68	0.65
Israel	0.82	0.82	0.67	0.65
Italy	0.84	0.81	0.68	0.68
Japan	0.76	0.77	0.66	0.55
Korea	0.78	0.79	0.73	0.64
Luxembourg	0.86	0.84	0.70	0.71
Mexico	0.86	0.86	0.81	0.79
Netherlands	0.80	0.78	0.72	0.64
New Zealand	0.82	0.77	0.69	0.62
Norway	0.80	0.79	0.63	0.56
Poland	0.89	0.88	0.71	0.75
Portugal	0.88	0.86	0.77	0.78
Slovak Republic	0.85	0.83	0.73	0.71
Slovenia	0.84	0.83	0.66	0.70
Spain	0.84	0.83	0.67	0.69
Sweden	0.81	0.77	0.65	0.60
Switzerland	0.81	0.79	0.67	0.65
Turkey	0.83	0.85	0.78	0.75
United Kingdom	0.78	0.75	0.72	0.59
United States	0.83	0.82	0.73	0.70
Median	**0.82**	**0.81**	**0.69**	**0.65**

Table 16.93 **Factor loadings and internal consistency of ESCS 2012 in partner countries and economies**

	Factor loadings			
	HISEI	**PARED**	**HOMEPOS**	**Reliability**
Argentina	0.83	0.82	0.75	0.69
Brazil	0.83	0.83	0.78	0.73
Bulgaria	0.81	0.81	0.74	0.71
Colombia	0.80	0.82	0.77	0.71
Costa Rica	0.83	0.81	0.81	0.76
Croatia	0.84	0.81	0.68	0.68
Cyprus[1, 2]	0.84	0.81	0.68	0.67
Hong Kong-China	0.85	0.83	0.78	0.76
Indonesia	0.81	0.82	0.79	0.73
Jordan	0.83	0.83	0.70	0.68
Kazakhstan	0.76	0.79	0.70	0.51
Latvia	0.84	0.82	0.73	0.73
Liechtenstein	0.78	0.79	0.68	0.62
Lithuania	0.82	0.81	0.74	0.69
Macao-China	0.79	0.80	0.72	0.65
Malaysia	0.83	0.76	0.78	0.70
Montenegro	0.81	0.80	0.73	0.68
Peru	0.84	0.84	0.79	0.76
Qatar	0.79	0.82	0.45	0.44
Romania	0.82	0.77	0.77	0.69
Russian Federation	0.80	0.78	0.72	0.63
Serbia	0.83	0.83	0.69	0.69
Shanghai-China	0.83	0.84	0.80	0.76
Singapore	0.85	0.84	0.73	0.73
Chinese Taipei	0.83	0.82	0.72	0.69
Thailand	0.86	0.87	0.82	0.76
Tunisia	0.83	0.84	0.80	0.75
United Arab Emirates	0.80	0.82	0.50	0.47
Uruguay	0.86	0.84	0.78	0.76
Viet Nam	0.82	0.82	0.79	0.74
Median	**0.83**	**0.82**	**0.74**	**0.69**

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
2. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Trends in ESCS

As explained above, the ESCS in PISA 2012 consisted of three sub-components, the highest parental occupation (*HISEI*), the highest parental education expressed as years of schooling (*PARED*) and the index of home possessions (*HOMEPOS)* which comprised all items on the *WEALTH*, *CULTPOS* and *HEDRES* scales, as well as books in the home (ST28Q01) recoded into a four-level categorical variable (fewer than or equal to 25 books, 26-100 books, 101-500 books, and more than 500 books).

In order to enable a trends study, the ESCS was computed in such a way that the ESCS scores would be more comparable across cycles. The ESCS was computed for the current cycle and also recomputed for the earlier cycles using a similar methodology. As in PISA 2012 the occupational coding scheme involved in the process of forming HISEI changed from ISCO-88 to ISCO-08, the occupational codes for previous cycles were mapped from the former to the current scheme (see also Chapter 3). In order to make the *PARED* sub-component of *ESCS* comparable across cycles, similar *ISCED* to *PARED* mapping schemes were employed for all the cycles. These mappings to years of education can be found in Annex D. To make the *HOMEPOS* sub-component more comparable across cycles, the scale was constructed in two steps. In the first step, a calibration sample over the five cycles was used to estimate international parameters for all items used over the five cycles. Items that were not administered in a certain cycle were treated as structurally missing data. This enabled comparisons across countries for these scales, the relative positions of the countries being estimated on a joint scale. When WLEs were estimated in the second step, the international parameters were anchored but the parameters corresponding to the items specific to each country, namely ST26Q15 to ST26Q17 were not fixed and were estimated during this run. The PCA for obtaining ESCS scores was then calculated across all cycles using these three comparable sub components (*HISEI*, *PARED* and *HOMEPOS*). The common weights for the PCA across cycles can be seen in Table 16.94.

Table 16.94 ESCS component weights use across cycles 2000, 2003, 2006, 2009 and 2012

ESCS sub-component weights		
HISEI	PARED	HOMEPOS
0.79	0.82	0.74

Computation of Scores based on Anchoring Vignettes in the PISA 2012 Student Questionnaire

As discussed in Chapter 3 of this report, anchoring vignettes were used as one of the new item formats in PISA 2012 to address issues of cross-cultural comparability of responses to context questionnaires. In this section, details are presented about how responses to anchoring vignettes were used as an alternative way of scoring responses to Likert-type questionnaire items.

Table 16.95 lists the twelve anchored indexes included in the international database for 2012, namely *ANCATSCHL*, *ANCATTLNACT*, *ANCBELONG*, *ANCCLSMAN*, *ANCCOGACT*, *ANCINSTMOT*, *ANCINTMAT*, *ANCMATWKETH*, *ANCMTSUP*, *ANCSCMAT*, *ANCSTUDREL*, and *ANCSUBNORM*. It should be noted that these indices have the prefix "ANC" to indicate that they were anchored as these indices are also on the database in their unanchored form without that prefix. The reason that not all scaled indices from the Student Questionnaire were anchored is that due to the rotated administration, anchoring vignettes were in only two of the three questionnaire forms so that only the question on which indices were based in those forms could be anchored.

Table 16.95 Anchored indexes in the international database

Index	Index label
ANCATSCHL	Attitude towards School: Learning Outcomes (Anchored)
ANCATTLNACT	Attitude towards School: Learning Activities (Anchored)
ANCBELONG	Sense of Belonging to School (Anchored)
ANCCLSMAN	Mathematics Teacher's Classroom Management (Anchored)
ANCCOGACT	Cognitive Activation in Mathematics Lessons (Anchored)
ANCINSTMOT	Instrumental Motivation for Mathematics (Anchored)
ANCINTMAT	Mathematics Interest (Anchored)
ANCMATWKETH	Mathematics Work Ethic (Anchored)
ANCMTSUP	Mathematics Teacher's Support (Anchored)
ANCSCMAT	Mathematics Self-Concept (Anchored)
ANCSTUDREL	Teacher-Student Relations (Anchored)
ANCSUBNORM	Subjective Norms in Mathematics (Anchored)

Question 84 in the Student Questionnaire forms B and C contained the set of three vignettes (ST84Q01 to ST84Q03) used in the anchoring procedure captured three levels of classroom management that could be described as low, medium, and high. Students read the vignettes and indicated their level of agreement with the statements that the described teacher was in control of his or her classroom using the same 4-point agreement scale that was used for most questionnaire items in the Student Questionnaire. Depending on their rating standards and their interpretation of the four levels of the response scale, students could place the three vignettes on different response categories. For instance, one student might have "Agreed" that a teacher described in the vignette was in control of his/her classroom while another student might have "Strongly agreed" or "Disagreed" with this statement. Since the actual levels for the people described in the vignettes were invariant across respondents, the only reason answers to the vignettes would have differed across respondents was interpersonal incomparability.

Differences in the evaluation of the vignettes between students indicated that students differed with regard to how they interpreted the response scale, and that any comparisons based on raw item responses might have resulted in validity problems. The alternative scoring based on the vignettes proposed by Bertling et al. (forthcoming) addressed this problem: regardless of where on the 4-point response scale a student placed the vignettes, a student's self-report answer could be scored relative to the rating the of low, medium, and high vignettes as three anchors that were invariant across respondents.

Responses on the original 4-point rating scale were re-scaled into a 7-point scale representing all possible relative rank comparisons between the responses to a 4-point self-report Likert-type item in the Student Questionnaire and the responses to the vignettes. On this scale, the value of "1" represented a rating of the self-report item lower than the rating of the low vignette, the value "2" represented a rating of the self-report item at the level of the rating of the low vignette, the value "3" represented a rating of the self-report item higher than the rating of the low vignette but lower than the rating of the medium vignette and so forth. The maximum value, "7", was assigned where a student's response to a questionnaire item was higher than his or her rating of the high vignette. In other words, low values were assigned where a self-report rating was relatively low compared to the rating of the vignettes, and high values were assigned where a self-report rating was relatively high compared to the rating of the vignettes. In this way, the three vignettes were used to anchor student ratings, providing context for the ratings on other self-report questionnaire items that shared the same response scale. Scoring was applied at the individual level using each student's responses to the vignettes as an anchor for this student's self-report answers. Table 16.96 illustrates the differences in possible values assigned to the original and the anchored item responses.

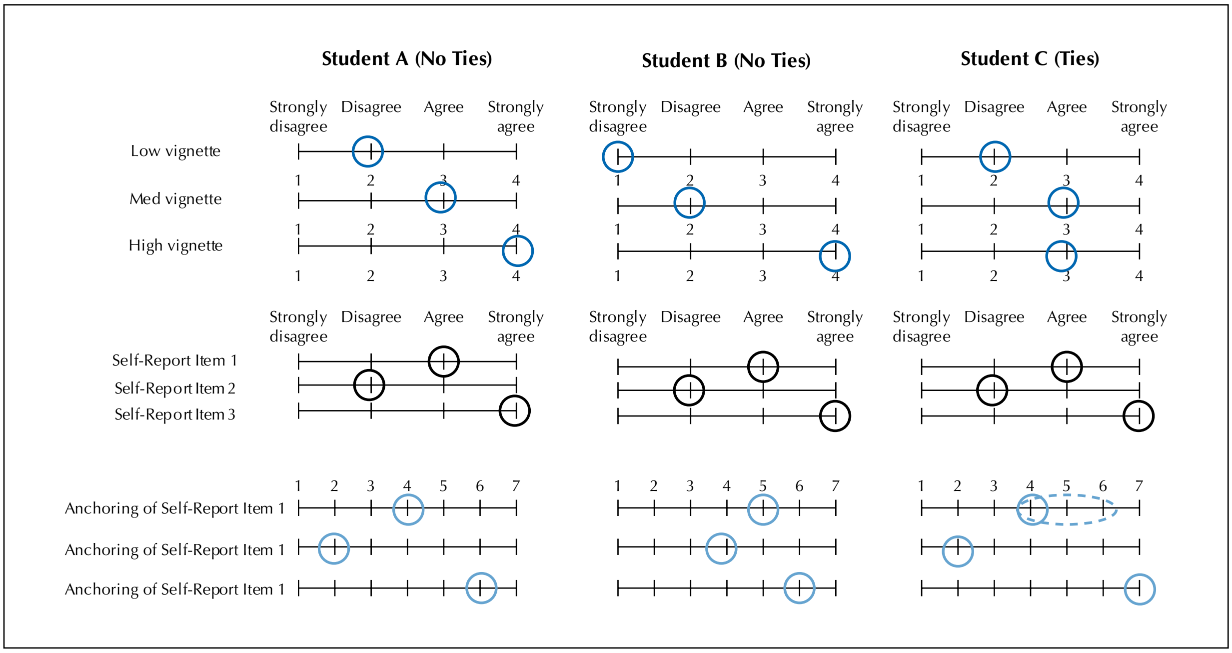
Table 16.96 Possible values for original and anchored item responses

Responses to a self-report item as presented in questionnaire	Strongly disagree	Disagree	Agree	Strongly Agree			
	1	2	3	4			
Anchored responses	**Lower than low vignette**	**Same as low vignette**	**In between low and medium vignette**	**Same as medium vignette**	**In between medium and high vignette**	**Same as high vignette**	**Higher than high vignette**
	1	2	3	4	5	6	7

A graphical illustration of the scoring procedure based on vignettes for three hypothetical examples of students' responses is given in Figure 16.3. The three hypothetical students in this example provided exactly the same answer to three self-report items but differed in their responses to the vignettes. As a consequence, scores on the anchored items also differed between the three students.

■ Figure 16.3 ■

Illustration of scoring based on vignettes for three hypothetical students (figure from Bertling and Kyllonen, 2013, with permission of the authors)



Two special cases during the calculation of anchored scores occurred when ties were encountered in the responses to the anchoring vignettes or when responses to the anchoring vignettes violated the expected order of vignettes. Ties occurred where a student chose the same response category for two or all three vignettes. Order violation occurred where a student chose a higher response category for a vignette representing a low value on the underlying construct than for a vignette representing a high value on the underlying construct. The scoring method used in PISA 2012 addressed these two special cases in the following ways.

Scoring if ties in the vignette ratings occurred: If students chose the same response category (e.g., "Agree") for multiple vignettes, self-report answers were scored based on "lower bound scores". This meant that the lowest possible score among the range of possible scores was assigned. This score reflected the value on a latent continuum that the respondent clearly pertained (i.e., the minimum) rather than a higher value that the respondent may or may not pertain. For example, if a student assigned the same ratings to the teachers with low and medium levels of classroom management, scores were adjusted based on the lower level (see Bertling et al. forthcoming for more details). The decision to use "lower bound" scores was also recommended by the PISA 2012 Technical Advisory Group who discussed preliminary findings for anchoring based on both Field Trial and Main Survey data at two meetings in 2012 and 2013.

Scoring if order violations in the vignette ratings occurred: To use the information provided with the responses of students who violated the expected order when rating the three vignettes, order violations were re-classified into ties. In other words, where a student rated the highest vignette lower than the medium vignette, responses for this student were rescaled in a way that the ratings for the medium and high vignette were tied. For instance, the rank order "low, high, med" would be rescaled into "low, {med, high}", with the brackets indicating that the same rank was assigned to the medium and high vignettes. It should be noted that, in most cases order violations were rescaled into complete ties of all vignettes (i.e., "{low, medium, high}"). While ties could be created in several ways during this post-hoc process, in PISA 2012 it was decided to create ties at the highest response category chosen by the student. For instance, in the aforementioned example ("low, {med, high}") the tie was created at the value which the respondent assigned to the high vignette. Ties were then scored as described above.

Four possible approaches to treating order violations were compared using initial Main Survey data from 52 countries to investigate whether or not the inclusion of anchored scores for these students in the international database would be feasible. To this end, the approaches were compared against the most conservative treatment of order violations, namely, the exclusion of students with such a response pattern. The first three approaches treated order violations by re-classifying "not-permitted" vignette ratings as ties. The first approach created ties at the lowest possible value, the second approach created ties at the medium value, and the third approach created ties at the highest possible value. In the fourth approach, order violations were transformed based on the ratings chosen by an individual irrespective of item content (cf. Fischer et al., 2009). The rationale for this alternative was that respondents might have mistakenly understood a preconceived "high" exemplar as a low one or a "low" exemplar as a high one. At the same time, the respondent might have been aware that exemplars varied in their standing on the dimension on which they were to be rated. Thus, through the pattern of ratings, respondents expressed their use of the rating scale (e.g., extreme, narrow, biased towards high, biased towards low), even when the respondent's ordering of ratings did not conform to the preconceived ordering of ratings. Following this logic, a respondent's self-rating could still be interpreted with respect to the vignette ratings, and mapped according to the nonmetric remapping rules, even though the respondent did not interpret the vignettes in accordance with their preconceived categories.

Results for the four approaches were compared, particularly in terms of potential differences in the correlations between indices and mathematics proficiency (see Chapter 17 for more details). Results indicated considerable variation in achievement correlations depending on the method used whereby, on average, the third approach showed the largest correlations, followed by the fourth approach. The details of the third approach whereby order violations were recoded into ties at the respondent's highest rating are provided in Table 16. 97. Detailed results regarding the comparison of different re-scaling approaches are described in Bertling, Kyllonen, Roberts, and Blew (forthcoming).

Table 16.97 Re-scaling rules for order violations on vignettes

Raw responses to vignettes			Recoding of actual values: order violations are recoded into ties of highest raw rating		
low	**medium**	**high**	**low**	**medium**	**high**
4	2	1	4	4	4
4	4	2	4	4	4
3	4	2	4	4	4
3	2	1	3	3	3
2	4	2	4	4	4
2	2	1	2	2	2
4	1	4	4	4	4
3	3	2	3	3	3
3	1	4	4	4	4
4	4	3	4	4	4
4	3	2	4	4	4
2	1	2	2	2	2
4	1	3	4	4	4
2	3	2	3	3	3
4	3	1	4	4	4
2	1	3	3	3	3
1	3	2	1	3	3
2	1	4	4	4	4
3	4	3	4	4	4
1	4	2	1	4	4
3	2	2	3	3	3
4	2	2	4	4	4
3	1	3	3	3	3
3	3	1	3	3	3
2	4	3	2	4	4
3	1	2	3	3	3
1	4	3	1	4	4
4	1	2	4	4	4
4	3	3	4	4	4
4	2	4	4	4	4
4	4	1	4	4	4
4	2	3	4	4	4
4	1	1	4	4	4
1	2	1	2	2	2
4	3	4	4	4	4
2	3	1	3	3	3
3	4	1	4	4	4
3	2	3	3	3	3
3	2	4	4	4	4
3	1	1	3	3	3
2	1	1	2	2	2
1	3	1	3	3	3
2	4	1	4	4	4
1	4	1	4	4	4

Assumptions and Cautions

The alternative scoring approach for Likert-type items based on anchoring vignettes makes the frame of reference for scoring of questionnaires items more transparent and can thereby help to interpret students' answers across different countries and educational systems. Several assumptions underlying the use of anchoring vignettes in the context of PISA have to be noted and careful consideration needs to be applied when interpreting anchored indices or using them in secondary data analyses.

First, the scoring approach is based on two main identifying assumptions, namely "vignette equivalence" and "response consistency" (see e.g., Kapteyn et al., 2011). The vignette equivalence assumption posits that different respondents interpret the vignette scenario in the same way. In other words, all differences in the ratings of the vignettes should be attributed to how different respondents interpret and use the response scale, not to any differences in their interpretation of the vignette scenarios themselves. The response consistency assumption posits that respondents use the same standards when evaluating themselves and when providing a rating of the vignette scenarios.

Secondly, the original anchoring vignette method was developed to anchor stand-alone questions only while in the context of the PISA 2012 student questionnaire the anchoring vignette method was extended so that the same vignette scenario was applied to a larger set of items. This extension was possible because of the third assumption that an

individual's rating standards were invariant across different contexts whenever the same rating scale was used. This meant that students were expected to use a four-point Likert scale with the categories "Strongly disagree" to "Strongly agree" in the same way for the different items in the student questionnaire, whether they were items such as "I learn mathematics quickly" or items such as "My teacher helps students with their learning".

Thirdly, the scoring process anchoring student responses using vignette scenarios depends on the particular vignette scenarios – that is where on the continuum of the underlying construct the vignettes are located – and the number of vignettes used. Analyses of the PISA 2012 data suggest reasonable consistency of results across the two sets of vignettes (see Chapter 17 for details on the second set of vignettes). Still, further research would be needed to understand fully the effects of different vignette contexts and the way in which the validity of results may depend on the number of vignettes and scale points used. For instance, gains in validity might be larger for questions that capture similar constructs as the constructs described in the vignettes. Also, future trend analyses of anchored indices would require the inclusion of the same vignettes in future questionnaires.

Fourth, the order of vignettes and self-reports in the questionnaire may have an influence on the results. As Hopkins and King (2010) showed, administering vignettes first might have a priming effect that reduces inter-individual differences in interpretation of the response scale. In the PISA 2012 Student Questionnaire, this might have been less of an issue as some self-report questions using the four-point Likert scale were presented before the vignettes and others were administered after the vignettes.

Finally, in order to use data from all students including those with tied ratings of anchoring vignettes or "order violations" additional assumptions were made as described above. Future research is intended to add to the understanding of students' response processes.

It is recommended that anchored indices should be interpreted in addition to original indices, not as a replacement. Values on both the original questionnaire indices and on anchored indices could be influenced by students' systematic or unsystematic response behaviours. Therefore, when undertaking analyses, both original and anchored indices should be used and results compared to arrive at an informed picture of the effects of using either type of index.

Notes

1. It should be noted that in the result for the international item parameters later in this chapter some item deltas are disordered. Adams, Wu and Wilson (2012b) have shown that rather than being indicative of a problem, such disordered deltas indicate specific patterns of relative numbers of respondents in each category.

2. A similar approach was used in the IEA Civic Education Study (see Schulz, 2004).

3. Only one principal component with an eigenvalue greater than 1 was identified in each of the participating countries.

References

Adams, R.J., Wu, M.L., and **Wilson, M.R.** (2012a), *ACER ConQuest 3.0.*, ACER, Melbourne.

Adams, R. J., M. L. Wu, and **M. Wilson** (2012b), "The Rasch rating model and the disordered threshold controversy", *Educational and Psychological Measurement*, 72(4), pp. 547-573.

Bertling, J. P., P. C. Kyllonen, R. D. Roberts and **E. Blew** (forthcoming), *Anchoring Vignettes for Improved Cross-Cultural Comparability of Survey Responses in International Educational Large-Scale Assessments.*

Bertling, J. P. and **P.C. Kyllonen** (2013), "Using anchoring vignettes to detect and correct for response styles in PISA questionnaires", in Prenzel, M. (Chair), *The Attitudes-Achievement-Paradox: How to Interpret Correlational Patterns in Cross-Cultural Studies,* Symposium at the EARLI 2013, Munich, Germany.

Fischer, R., Fontaine, J., van de Vijver, F.J.R. and **van Hemert, D** (2009), "What is style and what is bias in cross-cultural comparisons? An examination of response styles in cross-cultural research", in Gari A. and K. Mylonas (eds.), *Quod Erat Demonstrandum: From Herodotus' Ethnographic Journeys to Cross-Cultural Research*, pp. 137-148, Pedio, Athens.

Ganzeboom, H. B.G. (2010), *Questions and Answers about ISEI-08*, *http://www.harryganzeboom.nl/isco08/qa-isei-08.htm,* last accessed 8 October 2012.

Hopkins, D. J. and **G. King** (2010), "Improving Anchoring Vignettes: Designing Surveys to Correct for Interpersonal Incomparability", *Public Opinion Quarterly,* Vol. 74, pp. 201-22.

International Labour Organization (ILO) (2007), Resolution concerning updating the International Standard Classification of Occupations, *http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf,* last accessed 19 December 2013.

Kapteyn, A., J. P. Smith, A. Van Soest and **H. Vonkova** (2011), *Anchoring Vignettes and Response Consistency*, RAND Working Paper WR-840.

Masters, G. N. and **B. D. Wright** (1997), "The Partial Credit Model", in Van der Linden W. J. and R. K. Hambleton (eds.), *Handbook of Modern Item Response Theory,* Springer, New York/Berlin/Heidelberg.

OECD (1999), *Classifying Educational Programmes: Manual for ISCED-97Implementation in OECD Countries*, OECD Publishing, Paris. *http://www.oecd.org/education/skills-beyond-school/1962350.pdf*

OECD (2005), *PISA 2003 Technical Report*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264010543-en

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris. *http://dx.doi.org/10.1787/9789264190511-en*

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

Spiezia, V. (2010), "Does Computer Use Increased Educational Achievements? Student-level Evidence from NSA", *OECD Journal: Economic Studies*, 2010 (1).

Schulz, W. (2004), "Mapping Student Score to Item Responses", in Schulz W. and H. Sibberns (eds.), *Civic Education Study,* IEA, Amsterdam.

Schulz, W. (2003), *Validating Questionnaire Constructs in International Studies: TwoExamples from PISA200*, paper presented at the Annual Meeting of the American Educational Research Association (AERA) in Chicago, 21-25 April.

Traynor, A. and **Raykov, T.** (2013), "Household possessions indices as wealth measures: a validity evaluation", *Comparative Education Review*, *57*(4), pp. 662-688.

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

17

Questionnaire Development Issues

This chapter is concerned with a number of development issues in relation to the questionnaires used in PISA 2012. These issues include:

- the use of alternative question formats in an effort to deal with artifactual influences resulting from differences in response styles, both between respondents within countries but more so between countries;
- the lessons learned from the transition to a computer-delivered school questionnaire;
- comparative analyses undertaken because of the need to change from ISCO-88 to ISCO-08; and
- the implications of using rotated questionnaires for approaches to analysis.

ALTERNATIVE ITEM FORMATS

As discussed in Chapter 3, one of the major challenges of an international study such as PISA is the cross-cultural validity and applicability of all instruments. For PISA 2012, four specific alternative methods where explored in both the Field Trial and Main Survey as approaches to correcting for the impact of culturally-derived response styles that might bias the results (Kyllonen, Lietz and Roberts, 2010). In this section we report the evidence from the Field Trial and Main Survey concerning these methods. The four methods explored where anchoring vignettes, signal detection de-biasing based on the over-claiming technique, forced choice items, and situational judgement tests.

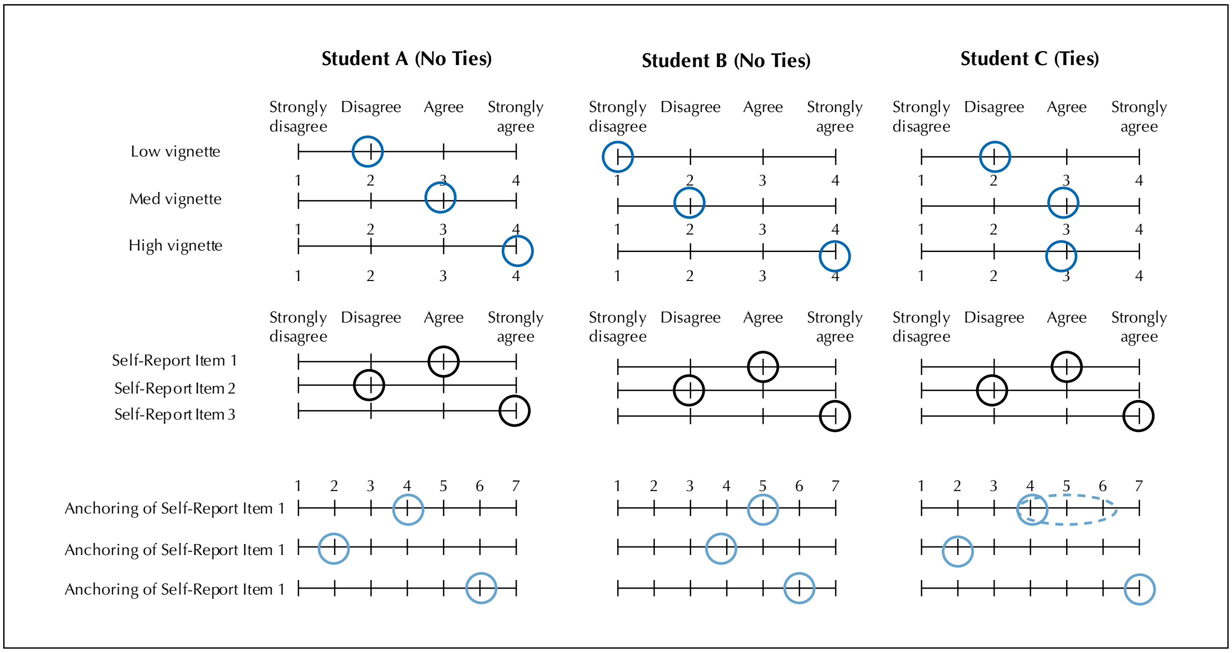
Anchoring vignettes

The original nonparametric anchoring vignette method was extended so that multiple items could be anchored based on the same set of anchoring vignettes (Bertling and Kyllonen, 2013). The scoring procedure is illustrated in the following, based on the classroom management vignettes.

As shown in Figure 17.1, the three vignettes capture three different levels of classroom management that can be described as low, medium and high (for actual wording of vignettes, see Chapter 3). Students were asked to read the vignettes and indicate their level of agreement with the statement that the described teacher is in control of his or her classroom using the same 4-point agreement scale that is also used for most questionnaire indices in the student questionnaire. Depending on their rating standards and their interpretation of the four levels of the agreement scale, students might place the three vignettes on different agreement categories. For instance, one student might "agree" that

■ Figure 17.1 ■

Illustration of scoring based on vignettes for three hypothetical students



Source: Bertling and Kyllonen (2013).

a teacher described in the first vignette is in control of his/her classroom while another student might “strongly agree” or “disagree” with this statement. Since the actual levels of teachers’ classroom management presented in the vignettes are invariant over respondents, differences in students’ response to the vignettes signal that students differ with regard to how they interpret the response scale, and that any comparisons based on raw item responses might have validity problems.

The alternative scoring based on the vignettes addresses this problem. When items were scored based on vignettes, numerical values for student responses were not assigned based on the concrete response option chosen (e.g., the value 4 for “strongly agree” and 3 for “agree”) but based on the self-report answer relative to the personal standard captured by the respondent’s individual rating of the three vignettes that form one set. Regardless of where on the 4-point agreement scale a student places the vignettes, a student’s self-report can be scored relative to his/her rating of low, medium and high for the vignettes. Based on this approach, in PISA 2012, students’ responses on the original 4-point agreement scale were re-scaled into a 7-point scale representing all possible relative rank comparisons of students’ self-reports and their rating of the vignettes. On this 7-point scale, the value 1 represents a rating lower than the low vignette, the value 2 represents a rating at the level of the low vignette, the value 3 represents a rating higher than the low but lower than the medium vignette, and so forth. The maximum score, 7, is assigned when a student’s self-reported response is higher than the rating of the high vignette. In other words, low values are assigned when a self-report rating is relatively low compared to the evaluation of the vignettes, and high values are assigned when a self-report rating is relatively high compared to the evaluation of the vignettes. In this way, the three vignettes are used to anchor student judgements, providing context for the ratings on other questions sharing the same response scale. Scoring is applied on the individual student level using each student’s responses to the vignettes as an anchor for this student’s self-reported responses to various Likert-type questions. Table 17.1 illustrates the differences in possible values assigned to original and anchored item responses.

Table 17.1 **Possible values for original and anchored item responses**

Responses to question as presented in questionnaire	**Strongly disagree**	**Disagree**	**Agree**	**Strongly agree**			
	1	2	3	4			
Anchored responses	**Lower than low vignette**	**Same as low vignette**	**In between low and medium vignette**	**Same as medium vignette**	**In between medium and high vignette**	**Same as high vignette**	**Higher than high vignette**
	1	2	3	4	5	6	7

A graphical illustration of the scoring procedure based on vignettes for three examples with and without ties is given in Figure 17.1. The three hypothetical students in this example provided exactly the same responses to the three self-reported items shown, but differ in their responses to the vignettes. As a result, scores on the anchored items also differ between the three students.

Clear interpretation of the vignettes in terms of the relative ordering of low, medium, and high levels of the described characteristics was one requirement for the use of vignettes. Two special cases are given when there are ties in the responses to the anchoring vignettes (i.e. a student chooses the same agreement category for two or all three vignettes) or when responses to the anchoring vignettes violate the expected order of vignettes (i.e. a student chooses a higher agreement category for a vignette representing a low value on the underlying construct than for a vignette representing a high value on the underlying construct). The scoring method used in PISA 2012 addresses these two cases in the following way.

Tied anchor evaluations

The anchoring vignette items administered in the student questionnaire used a 4-point Likert scale. Students had to place three vignettes on this scale. Not all students chose distinct values for each vignette, but assigned the same value to two or sometime all three vignettes instead. Several different approaches how to treat these cases with ties in the vignette rankings were investigated (Bertling and Kyllonen, 2013). A “lower bound” scoring seemed most promising based on these investigations. During scoring, “lower bound scores” were assigned to students with tied vignette responses. For example, if a student evaluated the teachers with low and medium levels of classroom management identical, scores were adjusted based on the lower level. Choosing the lower bound score created the largest variation and allowed for the best differentiation between individuals. Also, results show that using lower bound scores maximises the criterion-correlations.

Order violations in anchor evaluations

If students do not rate vignettes in the expected order, there are several options as to how these cases can be treated. Based on initial Main Survey data four different approaches to treat order violations were tested to investigate whether including adjusted scores for these students might be possible in the international database or not. That is, these methods were compared against the most conservative treatment of order violations, namely exclusion of students with order violations prior to anchoring adjustments.

Methods A-C: This approach treats order violations by re-classifying "unallowed" vignette evaluations as ties. That is, if a student rates the highest vignette lower than the medium vignette, responses for this student would be rescaled in a way that the ratings for the medium and high vignette are tied. For instance, the rank order "low, high, med" would be rescaled into "low, {med, high}", with the brackets indicating that the same rank is assigned to the medium and high vignette. Table 17.2 presents all possible patterns of order violations and the respective rescaled response patterns. Note that, in most cases order violations are rescaled into complete ties of all vignettes (i.e., "{low, medium, high}"). Whenever ties are created in this post-hoc manner, there are at least two or three different variants how these ties can be created. For instance, in the example used above ("low, {med, high}") the tie could be created at the value the respondent assigned to the medium, or to the high vignette. All possible variants of creating ties were implemented based on initial Main Survey data for 52 countries. Method A created ties at the lowest possible value, method B created ties at the medium value, and method C created ties at the highest possible value. Details are provided in Table 17.2.

Method D: This approach treats order violations by essentially ignoring the predesignated vignette categories. That is, method D involves treating the vignette with the highest rating by an individual as the "high" exemplar, the one with the lowest rating as the "low" exemplar, and the one in the middle as the "middle" exemplar regardless of their predesignated vignette assignments. If the individual rated the lowest predesignated vignette as the highest, method D treats it as the highest. Thus order violations are transformed into non-violations. The rationale for this approach is as follows. The response style literature identifies a common phenomenon characterised as a tendency for respondents to respond to a rating scale in a way "irrespective of item content" (Fischer et al., 2009). Given this phenomenon, it should not be surprising if due to inattention, comprehension failures, or focusing on the wrong vignette description features, a respondent might mistakenly understand a predesignated "high" exemplar as a low one or a "low" exemplar as a high one. At the same time, the individual might be aware that exemplars vary in their standing on the dimension on which they are to be rated. Thus through the pattern of ratings respondents express their use of the rating scale (e.g., extreme, narrow, biased towards high, biased towards low), even when the respondent's ordering of ratings does not conform to the predesignated ordering of ratings. Following this logic, a respondent's self-rating can still be interpreted with respect to the vignette ratings, and mapped according to the non-metric remapping rules, even though the respondent did not interpret the vignettes in accordance with their predesignated categories.

Results for methods A-D were compared based on initial Main Survey data, especially their impact on the correlations with mathematics proficiency. Findings indicated considerable variation in achievement correlations depending on which method is used, with method C resulting in, on average, the largest correlations, followed by method D. Following discussions with the PISA Technical Advisory Group (TAG), method C was chosen as an appropriate re-scaling method for the international database. Under this method, if order violations in the vignette ratings are present, they are re-classified into ties. That is, if a student rates the highest vignette lower than the medium vignette, responses for this student would be re-scaled in a way that the ratings for the medium and high vignette are tied. For instance, the rank order "low, high, med" would be rescaled into "low, {med, high}", with the brackets indicating that the same rank is assigned to the medium and high vignette. Note that, in most cases, order violations are rescaled into complete ties of all vignettes (i.e. "{low, medium, high}"). Ties are created at the highest response category chosen by the student. For instance, in the example used above ("low, {med, high}") the tie is created at the value the respondent assigned to the high vignette. Ties are then analysed as described above. Table 17.2 shows the rescaling procedure for order violations for all four possible methods.

Comparison of original and anchored indices

Field Trial analyses and initial Main Survey analyses consistently showed that within-country correlations with achievement tended to be higher for anchored scales, and correlations at the between-student within-country level and the between-country level did not show the inconsistencies found for unanchored scales. No "paradoxical" correlations were found for any anchored index, but correlations on the between-country level were similar to student-level correlations, both within countries and for the pooled sample. The absolute values of the between-country correlations tended to be larger than the correlations at the between-student within-country level.

Table 17.2 Detailed rescaling specifications for the different methods to address order violations

				Recoding into ties									Re-ordering		
				Method A			Method B			Method C			Method D		
				Recoding of actual values, tied at lower			Recoding of actual values, tied at medium			Recoding of actual values, tied at high			Re-ordering or original responses		
Original response	Valid %	Type	Recoding into allowed type with ties	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette
^421	0.04%	3,2,1	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^422	0.05%	3,(1,2)	(1,2,3)	2	2	2	4	4	4	4	4	4	2	4	4
^342	0.05%	3,1,2	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^342	0.05%	3,1,2	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^321	0.05%	3,2,1	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^242	0.07%	(1,3),2	(1,2,3)	2	2	2	2	2	2	4	4	4	2	2	4
^221	0.07%	3,(1,2)	(1,2,3)	1	1	1	2	2	2	2	2	2	1	2	2
^414	0.08%	2,(1,3)	(1,2,3)	1	1	1	4	4	4	4	4	4	1	4	4
^332	0.08%	3,(1,2)	(1,2,3)	2	2	2	3	3	3	3	3	3	2	3	3
^314	0.08%	2,1,3	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^443	0.09%	3,(1,2)	(1,2,3)	3	3	3	4	4	4	4	4	4	3	4	4
^432	0.09%	3,2,1	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^212	0.12%	2,(1,3)	(1,2,3)	1	1	1	2	2	2	2	2	2	1	2	2
^413	0.12%	2,3,1	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^232	0.13%	(1,3),2	(1,2,3)	2	2	2	2	2	2	3	3	3	2	2	3
^431	0.13%	3,2,1	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^213	0.14%	2,1,3	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^132	0.14%	1,3,2	1,(2,3)	1	2	2	1	3	3	1	3	3	1	2	3
^214	0.15%	2,1,3	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^343	0.15%	(1,3),2	(1,2,3)	3	3	3	3	3	3	4	4	4	3	3	4
^142	0.15%	1,3,2	1,(2,3)	1	2	2	1	4	4	1	4	4	1	2	4
^322	0.18%	(2,3),1	(1,2,3)	2	2	2	2	2	2	3	3	3	2	2	3
^422	0.20%	(2,3),1	(1,2,3)	2	2	2	2	2	2	4	4	4	2	2	4
^313	0.21%	2,(1,3)	(1,2,3)	1	1	1	3	3	3	3	3	3	1	3	3
^331	0.23%	3,(1,2)	(1,2,3)	1	1	1	3	3	3	3	3	3	1	3	3
^243	0.25%	1,3,2	1,(2,3)	2	3	3	2	4	4	2	4	4	2	3	4
^312	0.26%	2,3,1	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^143	0.41%	1,3,2	1,(2,3)	1	3	3	1	4	4	1	4	4	1	3	4
^412	0.42%	2,3,1	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^433	0.42%	(2,3),1	(1,2,3)	3	3	3	3	3	3	4	4	4	3	3	4
^424	0.47%	2,(1,3)	(1,2,3)	2	2	2	4	4	4	4	4	4	2	4	4
^441	0.50%	3,(1,2)	(1,2,3)	1	1	1	4	4	4	4	4	4	1	4	4
^423	0.52%	2,3,1	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^411	0.54%	(2,3),1	(1,2,3)	1	1	1	1	1	1	4	4	4	1	1	4
^121	0.54%	(1,3),2	(1,2,3)	1	1	1	1	1	1	2	2	2	1	1	2
^434	0.57%	2,(1,3)	(1,2,3)	3	3	3	4	4	4	4	4	4	3	4	4
^231	0.60%	3,1,2	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^341	0.79%	3,1,2	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^323	0.93%	2,(1,3)	(1,2,3)	2	2	2	3	3	3	3	3	3	2	3	3
^324	1.05%	2,1,3	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^311	1.55%	(2,3),1	(1,2,3)	1	1	1	1	1	1	3	3	3	1	1	3
^211	1.72%	(2,3),1	(1,2,3)	1	1	1	1	1	1	2	2	2	1	1	2
^131	3.03%	(1,3),2	(1,2,3)	1	1	1	1	1	1	3	3	3	1	1	3
^241	4.55%	3,1,2	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^141	5.64%	(1,3),2	(1,2,3)	1	1	1	1	1	1	4	4	4	1	1	4

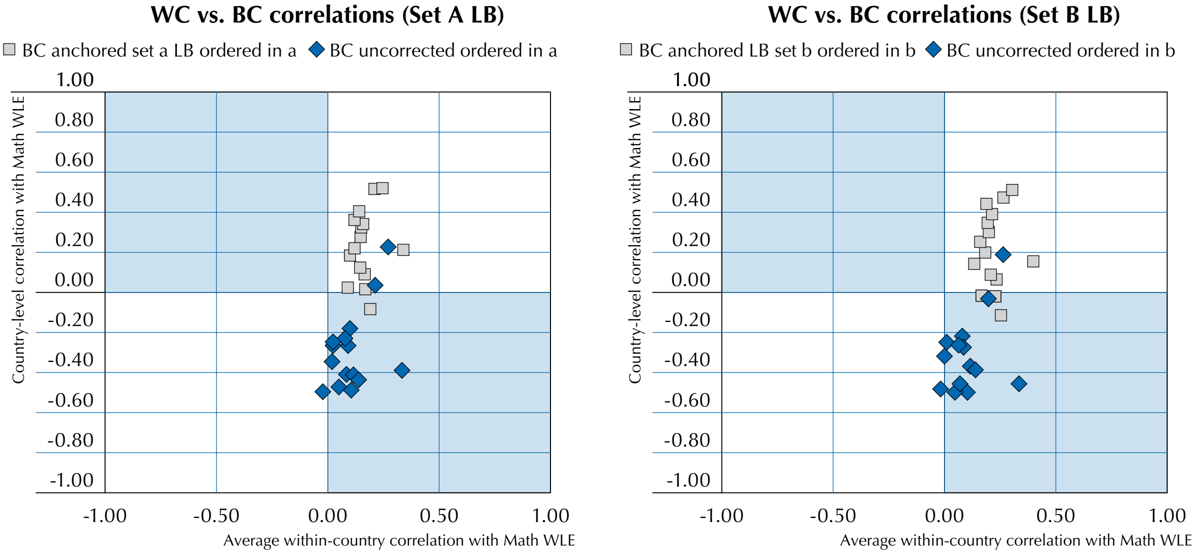
Note: Relative frequencies here are based on initial Main Survey data from 52 countries. Shaded fields indicate responses patterns that only partly violate the correct order of vignettes.

Comparisons of the two sets of anchoring vignettes showed very similar results with no major differences in the pattern of correlations between scales and achievement (see Figure 17.2 for a graphical illustration). Further, in comparison with the original indices, anchored indices showed smaller degrees of Differential Item Functioning (DIF) across countries and smaller correlations with indicators of acquiescence or disacquiescence response styles (Bertling and Kyllonen, 2013).

These findings supported the validity of the anchoring corrections and suggested that meaningful relationships of background indexes with achievement were suppressed by the dominance of general response tendencies among students when uncorrected responses were scaled.

■ Figure 17.2 ■

Alignment of average within-country and between-country correlations for unadjusted and anchored indexes



Note: In this figure, "LB" stands for lower bound and "Math WLE" stands for Mathematic: Weighted Least Square Estimate.

Four possible scenarios for making anchoring adjustments in the international database were presented to the PISA Technical Advisory Group (Bertling and Kyllonen, 2013). In scenario 1, only the two indices, namely Teacher Support (MTSUP) and Classroom Management (CLSMAN) would be anchored, as these scales directly map to the content of the vignettes. This scenario represented a construct specific anchoring approach. Scenarios 2-4 represent a construct non-specific approach. All questionnaire items with the same response format ("strongly agree" to "strongly disagree") can be anchored under these scenarios. Scenarios 2-4 differed in terms of which set of vignettes is used for the adjustment, or whether both sets are used. Results for the different scenarios were discussed with the PISA Technical Advisory Group (TAG). Anchoring adjustments for 12 Student Questionnaire scales were thereafter incorporated into the international database (see Chapter 16 for further details of all Student Questionnaire indices).

Limitations and cautions

The alternative scoring approach for Likert-type items based on vignettes makes the frame of reference for scoring of questionnaires items more transparent and can thereby help in interpreting students' responses across different countries and education systems. There are, however, several assumptions that underlie the use of anchoring vignettes in the context of PISA. Caution is advised when interpreting adjusted indices using anchoring vignettes (Bertling and Kyllonen, 2013).

The scoring approach is based on two main identifying assumptions: "vignette equivalence" and "response consistency" (e.g., Van Soest et al., 2011). First, the vignette equivalence assumption posits that different respondents interpret the vignette scenario in the same way. In other words, there is an assumption that all differences in the ratings of the vignettes should be attributable to the differences in how respondents interpret and use the agreement scale, but not to the differences in how respondents interpret the vignette scenario themselves. Second, the response consistency assumption posits that respondents use the same standards both in evaluating themselves and in providing an evaluation of the vignette scenario.

In this context, it should be noted that the original anchoring vignette method was developed to anchor items relating to the content described in the scenario of the anchoring vignette only. In the PISA 2012 Student Questionnaire, this method was extended so that the same vignette scenario was applied to a large set of different items. This extension was possible because of an assumption that an individual's rating standards are invariant across different contexts whenever the same rating scale is used. This means that students are expected to use a four-point Likert scale with the categories "Strongly disagree" to "Strongly agree" in a reasonably comparable way for the different questions included in the Student Questionnaire, whether these refer to items such as "I learn mathematics quickly" or items such as "My teacher helps students with their learning".

The scoring process anchoring student responses using vignette scenarios depends on the particular vignette scenarios (i.e. where on the continuum of the underlying construct the vignettes are located) and the number of vignettes used. While PISA 2012 data suggests reasonable consistency of results across the two sets of vignettes, further research is needed to fully understand the effects of different vignette contexts and how the validity of results depends on the number of vignettes and number of scale points used. For instance, gains in validity might be larger for questions that capture similar constructs as the constructs described in the vignettes.

The order of vignettes and self-reports in the questionnaire may have an influence on the results. As Hopkins and King (2010) showed, administering vignettes first might have a priming effect that reduces inter-individual differences in interpretation of the response scale. In the PISA 2012 Student Questionnaire some self-reported questions using the four-point Likert scale are presented before the vignettes and others are asked after the vignettes.

Finally, in order to use data from all students, including students with tied anchor evaluations (e.g. students who give the same ratings for two vignettes classified as low and medium) or "order violations" (e.g. students who give lower ratings to a vignette classified as high as to a vignette classified as medium or low), additional assumptions are needed, as described in the previous sections. Future research is needed to fully understand students' response processes.

It is recommended that adjusted indices using anchoring vignettes should be interpreted in addition to classical indices, not as a replacement. Both values on classical questionnaire indices and on adjusted indices can be influenced by students' systematic or unsystematic response behaviours. Examining both of these indices provides a basis for a more general picture of relationships and effects that is less tied to a single survey method only.

Topic familiarity with signal detection correction

Topic familiarity with signal detection correction of "Overclaiming Technique" (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy and Underwood, 1977), was used in the PISA 2012 as one way to enable adjustments for differences in response tendencies (see Chapter 3).

In the PISA 2012 Student Questionnaire an OCT was operationalised by asking students to indicate their familiarity – on a 5-point scale from "never heard of it" to "know it well; understand the concept" – with actual mathematics concepts (e.g. "polynomial function") and foils (e.g. "proper number"). Foils were created by combining a term from grammar (i.e. proper, as in proper noun; subjunctive, as in subjunctive mood; declarative as in declarative sentence) with a mathematical term (i.e. number; scaling; fraction, respectively). The questionnaire items are displayed in Figure 17.3.

Two indexes were computed from students' responses. One was a simple mean of their familiarity scores on the 5-point scale (FAMCON), and the other took that mean and subtracted from it the mean familiarity score of the foil concepts (FAMCONC).

Field Trial and initial Main Survey analyses (Bertling and Roberts, 2011) showed increased cross-cultural comparability of correlations with achievement for the adjusted compared to the unadjusted index. Also, familiarity ratings for the foil concepts only were strongly negatively correlated with achievement. Based on consultation with the PISA 2012 Technical Advisory Group who discussed preliminary findings from both Field Trial and Main Survey data at two meetings in 2012 and 2013, the adjusted topic familiarity index (FAMCONC) was included in the international database in addition to the original unadjusted index. While the overclaiming correction was applied to one index only in the international database, this correction might be applicable to other indices as well to correct for general response styles (Kyllonen, Bertling and Roberts, 2013).

Situational Judgement Tests

As discussed in Chapter 3, Situational Judgement Test items (SJTs; see Weekley and Ployhart, 2006, for an overview) were included in the PISA 2012 questionnaire as another possible way to adjust for differences in response behaviour.

The problem solving SJT in the PISA 2012 Student Questionnaire consisted of three different scenarios that described situations that could arise in the course of solving a problem. Questions focus on a person's initial response to a problem as well as possible approaches to take if one's initial response to the problem fails. The three scenarios involved a) a problem with a text message on a mobile phone, b) route selection for getting to a zoo and c) a malfunctioning ticket vending machine. Response options to each scenario tapped into different problem-solving strategies, namely systematic strategies, unsystematic strategies and seeking help.

■ Figure 17.3 ■

Topic familiarity items in the PISA 2012 Student Questionnaire (ST62)

ST62

Q **Thinking about mathematical concepts: how familiar are you with the following terms?**

(Please tick only one box in each row.)

	Never heard of it	*Heard of it once or twice*	*Heard of it a few times*	*Heard of it often*	*Know it well, understand the concept*
a) Exponential Function	☐1	☐2	☐3	☐4	☐5
b) Divisor	☐1	☐2	☐3	☐4	☐5
c) Quadratic Function	☐1	☐2	☐3	☐4	☐5
d) <Proper Number>	☐1	☐2	☐3	☐4	☐5
e) Linear Equation	☐1	☐2	☐3	☐4	☐5
f) Vectors	☐1	☐2	☐3	☐4	☐5
g) Complex Number	☐1	☐2	☐3	☐4	☐5
h) Rational Number	☐1	☐2	☐3	☐4	☐5
i) Radicals	☐1	☐2	☐3	☐4	☐5
j) <Subjunctive Scaling>	☐1	☐2	☐3	☐4	☐5
k) Polygon	☐1	☐2	☐3	☐4	☐5
l) <Declarative Fraction>	☐1	☐2	☐3	☐4	☐5
m) Congruent Figure	☐1	☐2	☐3	☐4	☐5
n) Cosine	☐1	☐2	☐3	☐4	☐5
o) Arithmetic Mean	☐1	☐2	☐3	☐4	☐5
p) Probability	☐1	☐2	☐3	☐4	☐5

Scenarios and items were designed along the goals of PISA that the PISA 2012 problem solving assessment should be based on problem solving in authentic concrete contexts rather than with abstract tasks (OECD, 2013a). A four point response format with the options “I would definitely not do this”, “I would probably not do this”, “I would probably do this”, and “I would definitely do this” was chosen. This is consistent with other background items in the PISA Student Questionnaires, where a 4-point response scale is the most frequently applied scale. Figure 17.4 shows an example item.

Based on exploratory factor analyses as well as more complex multi-trait multi-method models, three dimensions were identified in the SJT data associated with (a) systematic problem-solving behaviours (i.e., behaviours resulting from an analysis of the problem situation and planful acting), (b) unsystematic problem solving behaviours (i.e., unplanful and not the result of a thorough analysis of the problem situation but rather impulsive behavioural tendencies that are unrelated to the specific needs of the situation), and (c) help seeking behaviours (i.e., tendencies to rely on others’ knowledge and expertise regarding how to solve the problem; see Kyllonen, Bertling and Roberts 2013,

for details). While the three-factor structure was found across most countries, reliabilities of the possible indices were low (Cronbach's $\alpha < .70$). This is a typical finding for situational judgement tests that has been described in the literature many times (e.g., Christian et al., 2010). It was, however, decided not to include these indices in the international database as the internal consistencies fell short of the standards for internal consistencies of indices used in PISA.

■ Figure 17.4 ■

Example situational judgement test item from the PISA 2012 student questionnaire (ST96)

ST96

Q *Suppose that you have been sending text messages from your mobile phone for several weeks. Today, however, you can't send text messages. You want to try to solve the problem.*

What would you do? For each suggestion, tick the option that best applies to you.

(Please tick only one box in each row.)

	I would definitely do this	*I would probably do this*	*I would probably not do this*	*I would definitely not do this*
a) I press every button possible to find out what is wrong.	☐ 1	☐ 2	☐ 3	☐ 4
b) I think about what might have caused the problem and what I can do to solve it.	☐ 1	☐ 2	☐ 3	☐ 4
c) I read the manual.	☐ 1	☐ 2	☐ 3	☐ 4
d) I ask a friend for help.	☐ 1	☐ 2	☐ 3	☐ 4

Forced Choice

The Forced-Choice (FC) item format was another way in which PISA 2012 sought to overcome measurement issues related to the use of traditional Likert-type response formats (see Chapter 3).

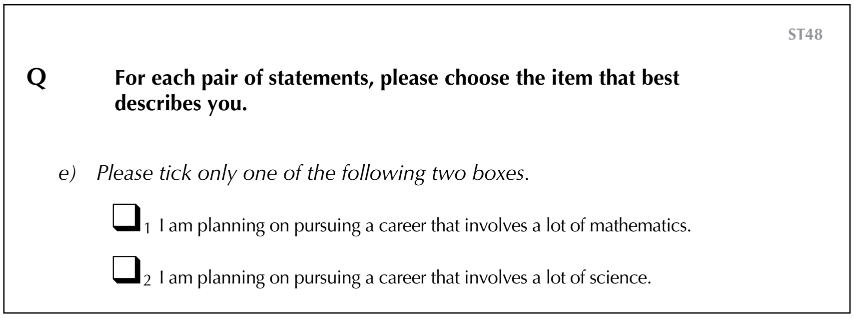
Five items used the forced-choice format to measure students' plans regarding mathematics at some stage in the future (*MATINTFC*) in PISA 2012. The first item had students decide between taking additional courses in mathematics or the language of the test after school finished. The second item asked whether students planned on majoring in a subject requiring mathematics or science at college (or equivalent educational institution in different countries). The third item asked whether students were willing to study harder in either their mathematics classes or in classes teaching the language of the test. For the fourth item, respondents had to indicate whether they were planning on taking as many mathematics or science classes as possible during their education. The fifth and last item of that battery required respondents to choose whether they were planning on pursuing a career that involved a lot of mathematics or science.

All items were reversed so that respondents who chose mathematics over either science or the test language were allocated a higher code. Then, responses to all five items were IRT (Item Response Theory) scaled and written onto the international database as a variable indicating mathematics intentions (*MATINTFC*).

One index based on the forced choice items is included in the international database.

■ Figure 17.5 ■

Example forced choice item from PISA 2012 Student Questionnaire (ST48e)



ST48

Q **For each pair of statements, please choose the item that best describes you.**

e) *Please tick only one of the following two boxes.*

☐1 I am planning on pursuing a career that involves a lot of mathematics.

☐2 I am planning on pursuing a career that involves a lot of science.

LESSONS CONCERNING ONLINE DELIVERY OF SCHOOL QUESTIONNAIRES

In addition to paper-based delivery of the questionnaires, PISA 2012 introduced online administration for the school questionnaire. On this first occasion online administration of the school questionnaire was optional.

Nineteen PISA participants took up the Online School Questionnaire option in the Main Survey in PISA 2012 which resulted in the administration of the questionnaire in 24 language versions. Participants included: Australia, Austria, Chile, Cyprus,[1] Denmark, Estonia, Finland, Hungary, Iceland, Ireland, Israel, Korea, Liechtenstein, Norway, Portugal, Singapore, Slovenia, Switzerland and Chinese Taipei.

Based on feedback from National Centres after the PISA 2012 Main Survey, the international contractor suggests that future cycles of PISA should consider how to further integrate the simultaneous processing of both paper-based and online questionnaires and materials during questionnaire development, negotiation, online authoring, verification, implementation and data management and cleaning.

While most processes were more efficient when compared to the paper-based administration, some issues specific to the Online School Questionnaire needed to be considered when interpreting results – particularly with regard to percentages of missing values for Online School Questionnaire data.

Questions SC14, SC16, SC18, SC22, SC25, SC33, SC34, and SC39 in the School Questionnaire contained more than six items. This required respondents to use a scroll-bar to view all items as the response scale for these items was fixed and had to always be visible. As such, higher levels of missing data for later items in 'long' questions were probable.

According to an analysis of Main Survey Online School Questionnaire data, on average, three to five per cent of respondents consistently failed to scroll-down for all later items, namely any item after the first six items, in these questions. Other missing values for items in these questions were randomly distributed, as respondents used the scroll-bar for some questions but failed to do so for others.

PISA participants that had more than five schools with missing values for all later items for SC14, SC16, SC18, SC22, SC25, SC33, SC34, and SC39 were given the opportunity to re-open their Online School Questionnaire to collect the missing data. Of the six participants that were approached two decided to re-open their questionnaires.

To address the question as to how the online administration mode compares to the paper-based administration mode, the average proportion of missing (code "99") across all questions in the School Questionnaire was computed for each country. Results are shown in Table 17.3.

Table 17.3 Average proportion of missing data across all questions in the School Questionnaire

CNT	online	% missing	CNT	online	% missing	CNT	online	% missing
ALB	no	2.91	GRC	no	3.18	NOR	yes	3.71
ARE	no	8.31	HKG	no	3.03	NZL	no	2.70
ARG	no	6.60	HRV	no	0.98	PER	no	4.60
AUS	yes	3.14	HUN	yes	1.62	POL	no	2.08
AUT	yes	5.76	IDN	no	3.99	PRT	yes	2.37
BEL	no	4.55	IRL	yes	4.43	QAT	no	5.02
BGR	no	4.03	ISL	yes	3.89	QCN	no	0.85
BRA	no	8.38	ISR	yes	3.53	ROU	no	0.22
CAN	no	3.55	ITA	no	4.67	RUS	no	1.57
CHE	yes	2.85	JOR	no	5.63	SGP	yes	1.84
CHL	yes	1.58	JPN	no	0.17	SRB	no	4.55
COL	no	7.02	KAZ	no	1.57	SVK	no	1.76
CRI	no	4.01	KOR	yes	1.09	SVN	yes	1.96
CYP*	yes	3.09	LIE	yes	3.17	SWE	no	2.79
CZE	no	3.60	LTU	no	2.46	TAP	yes	2.00
DEU	no	3.31	LUX	no	4.32	THA	no	0.81
DNK	yes	3.80	LVA	no	3.61	TUN	no	5.38
ESP	no	4.69	MAC	no	0.16	TUR	no	4.39
EST	yes	2.25	MEX	no	4.51	URY	no	1.20
FIN	yes	2.50	MNE	no	1.39	USA	no	4.91
FRA	no	4.36	MYS	no	4.39	VNM	no	0.71
GBR	no	2.32	NLD	no	2.77			

* See note 1 at the end of this chapter.

As can be seen in Table 17.3, the average proportion of missing data across all questions in the online mode ranged from 1.09% in Korea to 5.76% in Austria. For the paper-based School Questionnaire, average proportions of missing data ranged from 0.16% in Macao-China to 8.38% in Brazil. A comparison of the average proportion of missing data for online mode and paper-based mode of administrations revealed no substantive differences between the two modes.

Another way of exploring possible differences in missing data between administration modes is to look at individual items in the School Questionnaire. To this end, the items with the lowest and highest average proportion of missing data across countries were examined. Across both online and paper-based countries, the item with the lowest average proportion of missing data (0.17%) was recorded for the question that asked principals to indicate the size of the community in which their school was located (SC03Q01). For the online countries, the average proportion of missing data was 0.16% across the three online countries that had missing data for this question. For the paper-based countries, the average proportion of missing was very similar at 0.17% with ten of the 47 countries recording missing data. One of the highest average proportions of missing data across countries (18.9%) was recorded for SC10Q22 which asked about the number of part-time mathematics teachers with <ISCED 5A> qualification. The average proportion of missing data across the 19 countries which administered this question online was 5.62% while the average across the 46 countries which administered this version in paper-based mode was 24.89%. This difference indicated a considerably lower proportion of missing data for this item in the online mode compared with the paper-based mode (see Tables 17.13, 17.14 and 17.15 available at *www.oecd.org/pisa*).

Improvements in online administration of surveys for future cycles of PISA could consider the network response times for respondents that are geographically farther away from the server. With the server being located in Europe, participants outside this continent reported more problems with log-in and loading times for the questionnaire and between screens.

Lastly, future online questionnaire administration could consider how to improve questionnaire administration for respondents with widely differing local operating environments. Often, respondents were unable to update local browsers and operating systems due to centralised restrictions placed upon school electronic information systems. Future cycles of questionnaire administration should further consider how questionnaire and platform development as well as data management can accommodate the local operating environments of respondents.

THE EFFECT OF ISCO-08 COMPARED WITH ISCO-88 ON ESCS AND PERFORMANCE

In the PISA 2012 Main Survey, all questions relevant to the computation of ESCS (the *PISA index of economic, social and cultural status*) were included in the common part of the Student Questionnaire, which was administered to all students (see also the section on the rotated Student Questionnaire design Chapter 3).

In a first step, the International Standard Classification of Occupations in their 2008 version (ISCO-08) codes were mapped back onto the ISCO codes in their 1988 (ISCO-88) version. In addition, the most recent transformation from ISCO-08 to ISEI-08 published on Harry Ganzeboom's website *http://home.fsw.vu.nl/hbg.ganzeboom/isco08/*[2] was used.

ESCS was calculated using the resultant International Socio-Economic Index of occupational status (ISEI-08) codes in the same way as detailed in the *PISA 2006 Technical Report* (OECD, 2009), with straightforward changes due to differently named variables on the scale measuring home possessions (HOMEPOS). The computation for this comparison is slightly different from that used for the official database, so the results are not identical. In Table 17.4, the mean ESCS values using ISCO-88 and ISCO-08 are given for 64 of the 65 countries that participated in PISA 2012 as problems relating to the questionnaire data required for ESCS occurred in Albania. The first column shows the country. The next four columns show the means and standard deviations (S.D.) for the two ESCS versions using ISCO-88 (columns 2-3) and using ISCO-08 (columns 4-5). The next two columns report the absolute difference in the means between the two ESCS versions (column "Abs Diff") and the correlation between the two ESCS versions (column "Corr.").

It can be seen that the magnitude of the absolute differences is small with an average across all countries of 0.02, with 47 of the 64 countries in the analysis having a difference lower or equal to this value. The two largest absolute differences in the means are recorded for Jordan (0.07) and United Arab Emirates (0.06). The lowest correlations can be observed for Japan, Qatar and Spain with values of 0.94, 0.93 and 0.93 respectively. All other correlations are above 0.95.

As another way of examining the implication of the change in ISCO coding, effects of ESCS on student performance were investigated. In PISA 2012, the effect of ESCS on mathematics performance in a regression analysis ranged from 17 PISA points for Macao-China to 58 PISA points for Chinese Taipei (OECD, 2013b). This corresponds to a change in achievement of this number of PISA points for each unit change in ESCS. So, taking the first coefficient of 17 as an example, if a country has an average mathematics performance of 500 PISA points, one unit increase (=+1.0) in ESCS would correspond to an average performance of 517.

The average absolute change in mean ESCS for the countries is 0.02, which, depending on the size of the regression coefficient – from about 17 to 58 in mathematics achievement in PISA 2012 – would lead to a shift in the value of ESCS associated with a certain level of achievement. As the standard deviations are almost unchanged, only the intercept of the regression would change. For the two countries with the greatest change in ESCS, this would have the following implications. For Jordan, in mathematics in 2012, with a score-point difference associated with one unit on the ESCS of 22 the 0.07 change leads to about 22*0.07=1.6 PISA points change in the intercept, whereas the United Arab Emirates with a regression coefficient of 33 together with a change in ESCS of 0.06 would lead to a change of 33*0.06=2.0 PISA points.

Hence, the country ESCS means would seem not to have been greatly affected by the change in the ISCO and subsequent ISEI coding.

To understand better how the distribution of ESCS was affected, percentile plots of the two ESCS versions were performed. The circles in the plots below show the plot of the 1st, 2nd to 99th percentiles points of the two ESCS distributions. The fitted spline line is a piecewise defined low-order polynomial.

Figure 17.6 shows a percentile plot of the ESCS using the new version of ISEI-08 ("escsP") against the ESCS using ISEI-88 ("escs_88P") for Hong-Kong China as a case with a relatively higher correlation between the two measures using PISA 2012 data from the Main Survey. The results show very little nonlinearity.

Figure 17.7 shows the same plot for Qatar, which has a slightly lower, but still high, correlation between the two ESCS versions than other countries. A small amount of curvature can be observed at the top end but no serious indication of nonlinearity.

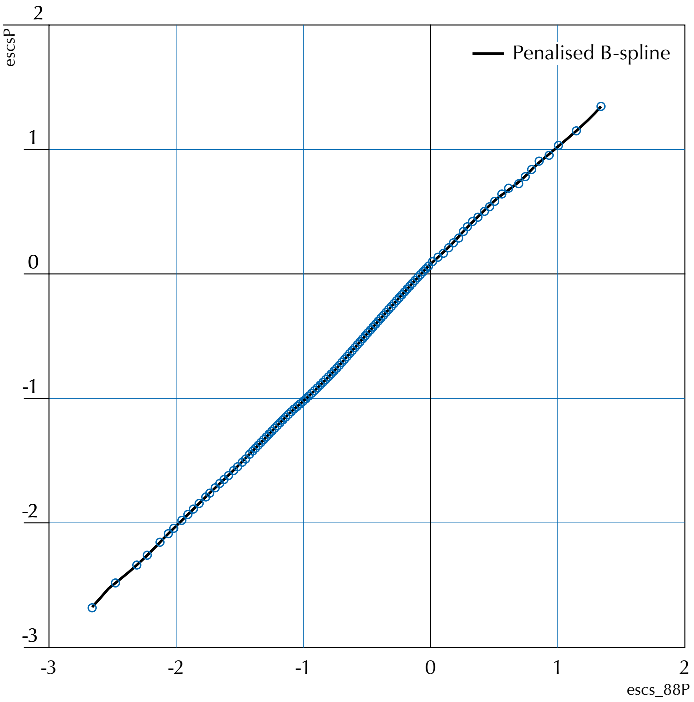
Table 17.4 ESCS calculated with alternative ISCO codings for PISA 2012 Main Survey

	ESCS using ISCO-88		ESCS using ISCO-08 (mapping)		Difference between the two ESCS versions	
CNT	Mean	SD	Mean	SD	Abs Diff	Corr.
ARE	0.40	0.87	0.45	0.84	0.06	0.95
ARG	-0.64	1.10	-0.62	1.10	0.02	0.97
AUS	0.23	0.80	0.23	0.79	0.00	0.95
AUT	0.12	0.85	0.10	0.86	0.02	0.97
BEL	0.15	0.91	0.15	0.91	0.00	0.97
BGR	-0.17	1.09	-0.15	1.04	0.02	0.97
BRA	-1.12	1.21	-1.13	1.19	0.02	0.98
CAN	0.41	0.86	0.41	0.86	0.00	0.97
CHE	0.19	0.90	0.17	0.89	0.02	0.97
CHL	-0.58	1.14	-0.58	1.14	0.00	0.98
COL	-1.16	1.18	-1.17	1.19	0.01	0.98
CRI	-0.90	1.22	-0.88	1.24	0.03	0.98
CYP*	0.21	0.89	0.19	0.89	0.01	0.97
CZE	-0.04	0.73	-0.05	0.73	0.01	0.96
DEU	0.24	0.93	0.21	0.93	0.02	0.97
DNK	0.39	0.82	0.43	0.83	0.03	0.96
ESP	-0.19	1.03	-0.19	1.03	0.00	0.93
EST	0.14	0.80	0.14	0.79	0.01	0.97
FIN	0.32	0.77	0.35	0.77	0.03	0.96
FRA	-0.04	0.80	-0.04	0.79	0.00	0.96
GBR	0.24	0.80	0.28	0.80	0.04	0.95
GRC	-0.05	0.98	-0.06	1.00	0.01	0.98
HKG	-0.68	0.93	-0.67	0.96	0.01	0.97
HRV	-0.18	0.84	-0.22	0.84	0.04	0.96
HUN	-0.22	0.94	-0.24	0.95	0.03	0.97
IDN	-1.73	1.07	-1.76	1.11	0.03	0.98
IRL	0.10	0.84	0.11	0.84	0.01	0.97
ISL	0.75	0.82	0.78	0.82	0.03	0.95
ISR	0.13	0.88	0.16	0.87	0.02	0.96
ITA	0.00	0.98	-0.02	0.96	0.02	0.97
JOR	-0.37	1.03	-0.30	1.01	0.07	0.96
JPN	-0.09	0.69	-0.12	0.69	0.03	0.94
KAZ	-0.22	0.78	-0.21	0.77	0.01	0.96
KOR	-0.08	0.72	-0.03	0.73	0.05	0.95
LIE	0.44	0.92	0.40	0.90	0.04	0.97
LTU	-0.01	0.93	0.00	0.91	0.01	0.97
LUX	0.10	1.12	0.11	1.09	0.01	0.98
LVA	-0.13	0.93	-0.14	0.89	0.01	0.97
MAC	-0.77	0.85	-0.77	0.86	0.00	0.97
MEX	-1.16	1.31	-1.16	1.29	0.00	0.99
MNE	-0.12	0.89	-0.14	0.90	0.01	0.95
MYS	-0.67	0.96	-0.63	1.00	0.05	0.97
NLD	0.24	0.78	0.23	0.77	0.01	0.96
NOR	0.44	0.79	0.42	0.76	0.02	0.96
NZL	0.01	0.83	0.03	0.82	0.02	0.95
PER	-1.13	1.25	-1.15	1.24	0.02	0.99
POL	-0.17	0.86	-0.18	0.88	0.01	0.97
PRT	-0.45	1.20	-0.45	1.19	0.00	0.98
QAT	0.50	0.90	0.52	0.91	0.02	0.93
QCN	-0.23	0.94	-0.23	0.95	0.00	0.98
ROU	-0.36	0.96	-0.36	0.95	0.01	0.97
RUS	-0.01	0.78	0.01	0.76	0.02	0.95
SGP	-0.16	0.91	-0.14	0.92	0.02	0.97
SRB	-0.20	0.93	-0.19	0.90	0.01	0.98
SVK	-0.12	0.92	-0.16	0.92	0.04	0.97
SVN	0.11	0.87	0.10	0.87	0.02	0.97
SWE	0.25	0.82	0.26	0.81	0.01	0.96
TAP	-0.28	0.83	-0.30	0.83	0.02	0.96
THA	-1.21	1.17	-1.25	1.19	0.03	0.98
TUN	-1.15	1.28	-1.12	1.31	0.02	0.98
TUR	-1.47	1.10	-1.48	1.12	0.01	0.98
URY	-0.77	1.14	-0.78	1.14	0.01	0.98
USA	0.15	0.98	0.16	0.98	0.01	0.96
VNM	-1.71	1.13	-1.76	1.18	0.05	0.99
Average	**-0.23**	**0.94**	**-0.22**	**0.94**	**0.02**	**0.97**

*See note 1 at the end of this chapter.

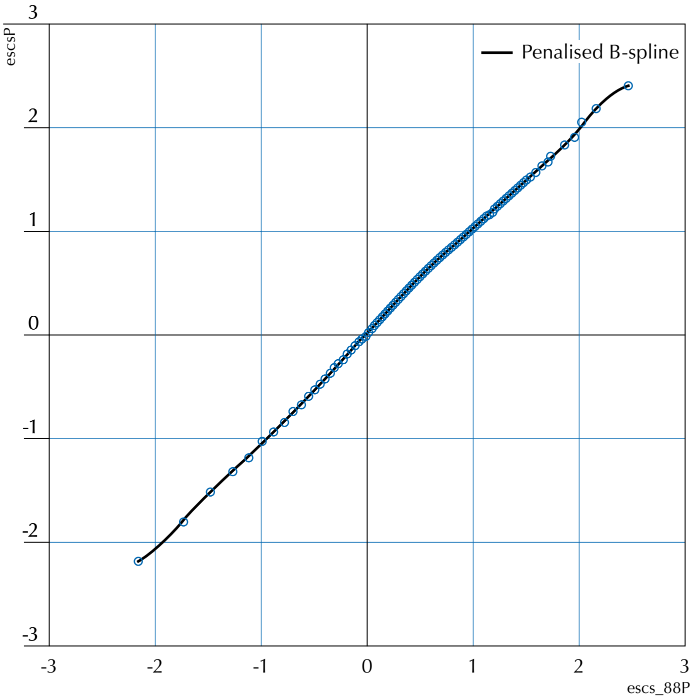
■ Figure 17.6 ■

Percentile plot of two ESCS versions – Hong Kong-China PISA 2012 Main Survey data



■ Figure 17.7 ■

Percentile plot of two ESCS versions – Qatar PISA 2012 Main Survey data



Percentile plots of the two ESCS versions were generated for all countries. Similar to Figure 17.7, plots seemed to show some flattening at the top end of the distributions for some countries, but, in general, the plots had little curvature as would be expected with the high correlations observed.

As the standard deviations for ESCS have not changed greatly in most countries, the effect of the update from ISCO-88 to ISCO-08 will be a shift of the intercept of the regression by up to 2 points of ESCS on achievement without changing the regression coefficient in most instances. Still, the analyses show a change in standard deviation of ESCS in a few countries. For example the standard deviation of ESCS in Indonesia increases from 1.07 to 1.11. This translates to a decrease in the coefficient by a factor of 1.07/1.11=0.96. In other words, instead of the regression coefficient taking on a value of 20 for ESCS in mathematics performance for Indonesia using the ISCO-08 coding the value would have been 21 using the ISCO-88 coding, which is not a large difference.

In the international database for PISA 2012, ESCS has been equated back to the earlier cycles using the 2012 ESCS as the base to establish trends rather than attempting to equate back to PISA 2009. This required the recomputing of ESCS using new ISCO-08 to ISEI-08 mapping for all cycles. Further details regarding the computation of ESCS are provided in Chapter 16.

An examination of the application of the new ISCO-08 code in the field

An analysis of the double coded mother's and father's occupation data from the Student Questionnaire administered as part of the PISA 2012 Field Trial (FT12) was conducted to investigate whether or not the change to the classification has impacted on the degree of agreement between the two coders.

In the FT12, two independent coders assigned a four digit ISCO code to mother's and father's occupation for a randomly selected number of students in 18 countries. A data file was produced with this information totalling 8,519 cases. Table 17.5 below shows the number of cases that were double coded for each country.

Table 17.5 Cases in the analysis of the PISA 2012 Field Trial double coding

CNT	Frequency	Percentage
BRA	321	3.8
CHE	94	1.1
CHL	1086	12.7
COL	152	1.8
DNK	224	2.6
HRV	1092	12.8
ISL	61	0.7
ISR	397	4.7
MAC	2609	30.6
MEX	300	3.5
PRT	451	5.3
RUS	227	2.7
SGP	222	2.6
SVN	292	3.4
SWE	187	2.2
TAP	263	3.1
TUR	342	4.0
URY	199	2.3
Total	**8519**	**100.0**

Cases were selected where the ISCO codes for the four-digit comparison differed between the two coders whereby missing responses for one or both coders were excluded. Differences were observed for (355) 4.2% of cases for mother's occupation and (427) 5.0% of cases for father's occupation. These data were tabulated to review which ISCO codes might be causing discrepancies between coders.

As getting the first two of the four digit codes correct achieves 90% accuracy (see Ganzeboom, 2010), cases that matched on the first two codes were considered to coincide sufficiently and removed. This left 2.4% of mother coded occupations and 3.3% of father coded occupations with discrepancies that were further examined.

From an examination of this small number of cases, no obvious pattern emerged for particular occupational groups (i.e. codes). There were similar numbers of discrepancies (when ordered ascending by the first coder) for each major group. For first coders with the first digit of ‘0’ there was 0.8% of discrepant data with the second coder, ‘1’=18%, ‘2’=18%, ‘3’=13%, ‘4’=6%, ‘5’=11%, ‘6’=0.8%, ‘7’=8%, ‘8’=5%, ‘9’(valid)=9% and ‘9’ (PISA)=11%. Note that ‘0’ was the only new group for ISCO-08.

The pattern of discrepancies seemed related more to the National Centre/country than to the ISCO code groupings as demonstrated by variation in the distribution of discrepant cases between countries. The discrepant data included cases from CHE (1.2%), COL (3.3%), DNK (17.4%), HRV (7.2%), ISL (2.1%), ISR (45.2%), MEX (12.4%), RUS (3.5%), SVN (0.2%), TUR (0.2%), URG (1.0%) and SWE (5.8%). ISR accounted for the largest proportion (almost half) of discrepancies between coders.

For comparison, 2009 data were extracted to examine if the amount of discrepancy in the FT12 is comparable to when coders where using ISCO-88 classification. Five of the countries that participated in the FT12 double coding exercise also participated in the Main Survey 2009; however difficulties were encountered when attempting to access the original data for Chinese Taipei which left Chile, Denmark, Iceland and Portugal for the analyses.

It should be noted that in 2009 the number of double coded responses was more than 32,000 cases from four countries, namely Chile (11 178), Denmark (7 869), Iceland (435) and Portugal (12 572). As for the FT12 data, only those cases that had discrepant data between the two coders were selected. This consisted of 6198 cases (8% of mother’s occupation codes and 11% of father’s occupation codes with missing/invalid data removed). Filtering out those cases that matched on the first two digits, only 5% of mother and 7% of father occupation codes differed.

Double coding of occupational data appears to reveal a high level of consistency between coders. The data from the FT12 and Main Survey 2009 (MS09) suggest that ISCO-08 could be contributing to an improved accuracy. There did not appear to be any obvious issue with the coding of particular occupational groups but issues may be more associated with local training or understanding of ISCO coding in general – but this does not appear to be specific to ISCO-08 and perhaps equally existed in previous cycles of PISA. Considering the small proportion of discrepant data, it is suggested that the change in classification will not affect the integrity of the ISCO data and may even improve it.

QUESTIONNAIRE ROTATION AND IMPLICATIONS FOR THE INTERNATIONAL DATABASE AND DATA ANALYSIS

The structure of the international database for the data obtained from the student questionnaire is as follows. Questions in the common part contain responses from all students.

For questions in the rotated part, responses from two thirds of students are available. For each student, the international database contains data on all variables whereby missing are recorded with a ‘7’ for those questions that were not administered to a student to indicate that this missing is a result of the rotation. As a consequence, each student will have all variables coded: the ones that s/he has answered reflecting his or her responses and the ones that s/he was not administered showing a ‘7’. So, in the rotated part, each student record contains actual responses to two thirds of the questions while one third is shown as ‘7’, not administered due to the rotation design. Consistent with their use in the whole database, the non-response values (8) and (9), can, of course, still occur to indicate invalid (8) or missing (9) values where students left a question that they were administered unanswered.

For illustrative purposes a screenshot of selected variables from the student questionnaire file of the PISA 2012 international database is presented in Table 17.6. As it would have been too large to show all Student Questionnaire questions, this particular screenshot was taken of selected variables from the common and rotated part of the Student Questionnaire using actual data for the first twenty students in the data file for Albania. The variables in the file are as follows:

- CNT, SCHOOLID, STIDStd: Country code, 3 digit character; School ID, 7 digit; Student ID;
- ST01Q01 and ST04Q01: Questions from the common part (i.e. “International Grade” and “Gender”); hence valid codes for all cases;

- ST29Q01 and ST35Q01: Questions from the question sets included in Forms A and B (i.e. "Maths Interest - Enjoy Reading" and "Subjective Norms - Friends Do Well in Mathematics"); hence the cases with QUESTID=3 show the value of "7";
- ST53Q01 and ST55Q01: Questions from the question sets included in Forms A and C (i.e. " Learning Strategies - Important Parts vs. Existing Knowledge vs. Learn by Heart" and " Out of school lessons - <test lang>"); hence the cases with QUESTID=2 show the value of "7";
- ST77Q01 and ST79Q01: Questions from the question sets in Forms B and C (i.e."Math Teaching - Teacher Shows Interest" and "Teacher-Directed Instruction - Sets Clear Goals"); hence the cases with QUESTID=1 show the value of "7"; and
- QUESTID: Indicates which Student Questionnaire Form is used: 1=Form A, 2=Form B, 3=Form C; plus 5=Form UH (one hour, see Chapter 2 for more details) which contained the common part only (see Chapter 3 for more details on Student Questionnaire Forms. The questionnaire forms are available on *www.oecd.org/pisa*).

Therefore, going from right to left in Table 17.6 below, the data show that students 2 and 3 in the first school were administered Student Questionnaire Form A as indicated by the "1" for QUESTID in the last column. The two columns preceding this last column contain information on the variables ST77Q01 ("Math Teaching - Teacher Shows Interest") and ST79Q01 ("Teacher-Directed Instruction - Sets Clear Goals"). As these variables were in question set 3 which was not included in Student Questionnaire Form A, the code "7" indicates missing data due to these questions not having been administered to students 2 and 3.

Table 17.6 Selected variables in the student questionnaire file of the PISA 2012 database

CNT	SCHOOLID	STIDStd	ST01Q01	ST04Q01	ST29Q01	ST35Q01	ST53Q01	ST55Q01	ST77Q01	ST79Q01	QUESTID
ALB	0000001	00001	10	1	2	3	7	7	1	4	2
ALB	0000001	00002	10	1	3	1	2	1	7	7	1
ALB	0000001	00003	9	1	2	1	1	2	7	7	1
ALB	0000001	00004	9	1	7	7	2	1	1	9	3
ALB	0000001	00005	9	1	3	1	7	7	1	3	2
ALB	0000001	00006	9	1	7	7	2	1	1	1	3
ALB	0000001	00007	10	1	7	7	3	2	1	1	3
ALB	0000001	00008	10	2	2	2	7	7	1	1	2
ALB	0000001	00009	9	1	2	2	7	7	4	1	2
ALB	0000001	00010	10	1	7	7	2	1	2	3	3
ALB	0000002	00011	10	1	7	7	2	1	1	2	3
ALB	0000002	00012	10	1	2	2	2	1	7	7	1
ALB	0000002	00013	10	1	1	1	7	7	9	3	2
ALB	0000002	00014	10	1	2	1	1	4	7	7	1
ALB	0000002	00015	10	2	2	2	1	3	7	7	1
ALB	0000002	00016	10	2	1	2	7	7	1	1	2
ALB	0000002	00017	10	1	1	1	7	7	1	1	2
ALB	0000002	00018	10	2	3	9	3	1	7	7	1
ALB	0000002	00019	10	1	7	7	2	1	1	1	3
ALB	0000002	00020	10	1	7	7	1	2	1	1	3

The next four columns to the left show valid codes for students 2 and 3 who responded to variables in question sets 1 and 2 which were included in Form A, namely ST29Q01, ST35Q01, ST53Q01 and ST55Q01. In contrast, student 1, who responded to Student Questionnaire Form B as indicated by the "2" in QUESTID, has missing data due to the non-administration (i.e. code "7") recorded for the variables ST53Q01 and ST55Q01. Further, student 4, who responded to Student Questionnaire Form C as indicated by the "3" in QUESTID, has missing data due to the non-administration (i.e. code "7") recorded for the variables ST29Q01 and ST35Q01. It should also be noted that information on ST79Q01 is missing for student 4 despite the fact that this question was administered to this student. This is indicated by the code "9".

The two variables ST01Q01 and ST04Q01 contain valid codes for all students as the corresponding questions were included in the common part of the student questionnaire. While ST01Q01 shows that students 1 and 2 are in grade 10 and students 3 and 4 are in grade 9 all four students are female as indicated by the "1" in ST04Q01.

Implications for further analyses

The implications of the rotated Student Questionnaire design for further analyses differ depending on:

a. Whether the variables are located in the common or rotated part;

b. The intention of the analysis (e.g. examination of effects of school level variables on student performance).

It should be noted that rotation has no implication where proportions are reported. It also has no implications on the computation of standard errors or the use of replicate weights which will still provide the correct estimates. Likewise, rotation has no implications for the way that the replicate weights need to be used. The sum of the weights will be an estimate of the population size for students who responded to these items.

Rotation does have implications where frequencies or raw counts are reported and these values are used to estimate the population value of a particular variable. In this case, to get population estimates, raw values have to be multiplied by the inverse of the proportion who responded to the question. For questions from one question set to which 2/3 of students responded, values would have to be multiplied by a factor of 1.5.

Questions in the common part

Analyses using data from questions in the common part (e.g. the component questions constituting ESCS as well as gender, immigrant status and language at home) will be the same as in previous cycles as all students responded to these questions.

Details regarding the questions in the common part are given in Chapter 3.

Questions in the rotated part

While details regarding the questions in the rotated part of the Student Questionnaire are given in Chapter 3, this section provides further information regarding the implications of the rotated design on a) the set up of the international database, b) standard errors, c) effective sample sizes, d) intraclass correlations (rhos) of constructs and e) correlations between constructs in the Student Questionnaire and mathematics performance.

Due to the design of the Student Questionnaire, variables in the rotated part contain responses from two thirds of students in the sample. Thus, if a country has an average sample size of 30 students within a school, responses are obtained from 20 students. As an example, the analysis could include the constructs of Instrumental Motivation and Mathematics Interest (ST29Q01) and Subjective Norms (ST35Q01).

It should be noted that accuracy has been reduced slightly for variables in the rotated parts as the reduced sample size has increased the standard errors of the population estimates of the rotated variables. The increase in standard error for countries with a low intraclass correlation will be greater compared to countries with a high intraclass correlation. The range of the increase in standard errors of means is given in Table 17.7 while the effective sample sizes for different rotation designs are presented in Table 17.8.

Table 17.7 Range of the increase of the standard error of the means

	intraclass correlation, rho									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Rotated Student Questionnaire design PISA 2012	1.22	1.05	1.03	1.02	1.01	1.01	1.00	1.00	1.00	1.00

Table 17.8 Effective sample sizes for different rotation designs

	intraclass correlation, rho									
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
No rotation	5250	1193	673	469	360	292	245	212	186	166
Rotated Student Questionnaire design PISA 2012	3500	1082	640	455	352	288	243	210	186	166

As can be seen in Table 17.7, rotation has little impact on standard errors for higher rhos. Correspondingly, Table 17.8 shows that the impact of the rotation in terms of decrease to effective sample size declines as rho increases.

Intraclass correlations for mathematics achievement (MAch) and all questionnaire indices for OECD countries are included in Table 17.9 for PISA 2003 and Table 17.10 for PISA 2012. As can be seen, the average intraclass correlation for questionnaire scales across OECD countries in PISA 2003 was 0.07 which meant that the increase in the standard error of the means for these scales was between 1.22 and 1.05 (see Table 17.7). Correspondingly for PISA 2012, the average intraclass correlation for questionnaire scales across OECD countries was 0.10 which meant an average increase of 1.05 in the standard error of the means for these scales (Table 17.7).

[Part 1/2]

Table 17.9 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2003

CNT	MAch	ANXMAT	ATSCHL	ATTCOMP	BELONG	COMPHOME	COMPLRN	COOPLRN	CSTRAT	CULTPOS	DISCLIM	ELAB	ESCS	HEDRES
AUS	0.21	0.03	0.04	0.03	0.03	0.07	0.04	0.02	0.03	0.11	0.08	0.02	0.26	0.05
AUT	0.55	0.06	0.03	0.05	0.03	0.12	0.08	0.03	0.02	0.17	0.17	0.07	0.32	0.07
BEL	0.56	0.03	0.05	0.03	0.03	0.21	0.05	0.02	0.07	0.14	0.12	0.06	0.32	0.17
CAN	0.17	0.03	0.05	0.03	0.03	0.09	0.04	0.03	0.05	0.09	0.09	0.03	0.19	0.04
CHE	0.33	0.04	0.04	0.03	0.05	0.06	0.06	0.04	0.03	0.08	0.11	0.03	0.19	0.07
CZE	0.52	0.07	0.05	0.05	0.05	0.11	0.05	0.04	0.04	0.12	0.20	0.04	0.30	0.09
DEU	0.58	0.02	0.03	0.02	0.01	0.09	0.06	0.02	0.02	0.12	0.11	0.04	0.31	0.08
DNK	0.13	0.04	0.04	0.02	0.03	0.05	0.04	0.03	0.03	0.11	0.11	0.02	0.19	0.06
ESP	0.20	0.04	0.05	n/a	0.05	0.11	0.04	0.03	0.03	0.10	0.13	0.03	0.25	0.04
FIN	0.05	0.02	0.02	0.02	0.02	0.03	0.01	0.02	0.01	0.05	0.10	0.00	0.11	0.02
FRA	0.46	0.02	0.04	n/a	0.01	0.12	0.05	0.03	0.03	0.15	0.10	0.04	0.29	0.07
GBR	0.23	0.04	0.04	0.05	0.03	0.06	0.03	0.04	0.04	0.13	0.09	0.03	0.19	0.05
GRC	0.36	0.04	0.06	0.00	0.02	0.14	0.01	0.01	0.02	0.16	0.08	0.01	0.29	0.18
HUN	0.59	0.06	0.05	0.06	0.04	0.23	0.04	0.01	0.01	0.27	0.19	0.03	0.44	0.20
IRL	0.17	0.03	0.03	0.00	0.00	0.08	0.04	0.02	0.02	0.07	0.08	0.02	0.21	0.05
ISL	0.04	0.01	0.04	0.02	0.01	0.02	0.01	0.03	0.02	0.04	0.22	0.02	0.17	0.01
ITA	0.53	0.03	0.06	0.07	0.04	0.14	0.10	0.05	0.07	0.15	0.18	0.08	0.30	0.07
JPN	0.54	0.03	0.01	0.04	0.04	0.11	0.05	0.04	0.04	0.15	0.27	0.03	0.27	0.14
KOR	0.42	0.02	0.02	0.02	0.04	0.08	0.14	0.08	0.14	0.13	0.09	0.06	0.30	0.12
LUX	0.32	0.02	0.06	n/a	0.03	0.07	0.02	0.04	0.02	0.11	0.04	0.04	0.24	0.02
MEX	0.39	0.05	0.08	0.05	0.07	0.30	0.04	0.05	0.06	0.17	0.10	0.06	0.34	0.15
NLD	0.63	0.04	0.05	n/a	0.04	0.10	0.05	0.03	0.04	0.12	0.10	0.05	0.23	0.12
NOR	0.07	0.03	0.05	n/a	0.01	0.02	0.02	0.03	0.01	0.06	0.09	0.02	0.12	0.02
NZL	0.18	0.04	0.03	0.04	0.01	0.09	0.06	0.01	0.03	0.06	0.06	0.05	0.17	0.06
POL	0.13	0.02	0.03	0.01	0.02	0.11	0.02	0.02	0.01	0.09	0.11	0.02	0.23	0.05
PRT	0.34	0.03	0.03	0.02	0.03	0.16	0.01	0.03	0.05	0.11	0.08	0.02	0.24	0.09
SVK	0.43	0.06	0.05	0.06	0.04	0.17	0.06	0.03	0.04	0.10	0.15	0.05	0.32	0.19
SWE	0.11	0.03	0.03	0.03	0.01	0.03	0.04	0.01	0.02	0.08	0.11	0.02	0.12	0.02
TUR	0.56	0.07	0.05	0.03	0.03	0.24	0.04	0.03	0.05	0.14	0.10	0.04	0.37	0.24
USA	0.26	0.03	0.03	0.01	n/a	0.12	0.04	0.03	0.03	0.10	0.08	0.05	0.23	0.06

Note: n/a indicates country not administering questions forming construct.

[Part 2/2]
Table 17.9 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2003

CNT	HIGHCONF	HOMEPOS	INSTMOT	INTCONF	INTMAT	INTUSE	MATHEFF	MEMOR	PRGUSE	ROUTCONF	SCMAT	STUREL	TEACHSUP	AVG*
AUS	0.03	0.17	0.02	0.05	0.04	0.03	0.08	0.02	0.03	0.02	0.03	0.06	0.05	0.05
AUT	0.09	0.22	0.17	0.11	0.09	0.08	0.19	0.03	0.10	0.15	0.06	0.10	0.14	0.11
BEL	0.04	0.24	0.06	0.13	0.05	0.05	0.14	0.03	0.07	0.11	0.02	0.07	0.07	0.09
CAN	0.05	0.12	0.05	0.06	0.05	0.05	0.06	0.03	0.09	0.05	0.04	0.07	0.06	0.06
CHE	0.04	0.12	0.07	0.05	0.03	0.03	0.12	0.04	0.04	0.06	0.03	0.13	0.09	0.06
CZE	0.10	0.18	0.10	0.09	0.06	0.06	0.22	0.06	0.07	0.12	0.05	0.08	0.15	0.10
DEU	0.03	0.20	0.04	0.04	0.04	0.02	0.12	0.04	0.04	0.06	0.01	0.07	0.09	0.07
DNK	0.02	0.13	0.03	0.03	0.05	0.03	0.05	0.02	0.03	0.03	0.03	0.12	0.08	0.06
ESP	n/a	0.15	0.04	n/a	0.04	n/a	0.07	0.02	n/a	n/a	0.05	0.10	0.11	0.07
FIN	0.02	0.05	0.01	0.01	0.02	0.01	0.03	0.01	0.02	0.01	0.02	0.05	0.07	0.03
FRA	n/a	0.23	0.03	n/a	0.04	n/a	0.15	0.03	n/a	n/a	0.05	0.07	0.08	0.08
GBR	0.11	0.16	0.04	0.07	0.04	0.07	0.09	0.05	0.14	0.09	0.04	0.06	0.06	0.07
GRC	0.06	0.28	0.03	0.06	0.03	0.04	0.10	0.01	0.04	0.09	0.05	0.10	0.08	0.08
HUN	0.09	0.37	0.04	0.16	0.05	0.07	0.22	0.02	0.07	0.21	0.05	0.11	0.09	0.12
IRL	0.02	0.13	0.03	0.05	0.02	0.04	0.07	0.01	0.02	0.03	0.03	0.04	0.07	0.05
ISL	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.01	0.02	0.02	0.02	0.10	0.12	0.04
ITA	0.06	0.21	0.11	0.10	0.11	0.06	0.15	0.04	0.08	0.12	0.05	0.11	0.13	0.10
JPN	0.05	0.19	0.09	0.07	0.06	0.04	0.26	0.03	0.11	0.09	0.03	0.07	0.08	0.09
KOR	0.02	0.21	0.09	0.03	0.08	0.05	0.21	0.05	0.03	0.06	0.10	0.06	0.04	0.09
LUX	n/a	0.16	0.04	n/a	0.03	n/a	0.08	0.02	n/a	n/a	0.01	0.07	0.07	0.06
MEX	0.17	0.28	0.05	0.28	0.10	0.14	0.07	0.06	0.08	0.28	0.06	0.09	0.09	0.12
NLD	n/a	0.18	0.03	n/a	0.04	n/a	0.09	0.04	n/a	n/a	0.04	0.07	0.06	0.07
NOR	n/a	0.06	0.03	n/a	0.03	n/a	0.04	0.01	n/a	n/a	0.02	0.08	0.07	0.04
NZL	0.04	0.14	0.02	0.02	0.07	0.05	0.06	0.03	0.06	0.02	0.02	0.05	0.04	0.05
POL	0.05	0.15	0.03	0.11	0.03	0.10	0.05	0.00	0.04	0.08	0.02	0.06	0.07	0.06
PRT	0.03	0.21	0.03	0.09	0.03	0.06	0.11	0.03	0.02	0.10	0.03	0.06	0.07	0.07
SVK	0.10	0.23	0.10	0.16	0.07	0.12	0.23	0.03	0.08	0.17	0.06	0.14	0.17	0.11
SWE	0.03	0.07	0.02	0.06	0.04	0.04	0.06	0.01	0.02	0.03	0.03	0.06	0.04	0.04
TUR	0.04	0.36	0.04	0.15	0.07	0.07	0.20	0.03	0.08	0.15	0.07	0.05	0.05	0.11
USA	0.04	0.17	0.03	0.07	0.05	0.04	0.06	0.04	0.05	0.06	0.04	0.05	0.06	0.06

*Average rho across all questionnaire scales in this country.

Note: n/a indicates country not administering questions forming construct.

[Part 1/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	Math	ANXMAT	ATSCHL	ATTLNACT	BELONG	CLSMAN	COGACT	CULTDIST	CULTPOS	DISCLIMA	ENTUSE	ESCS	EXAPPLM	EXPUREM
AUS	0.28	0.04	0.08	0.05	0.07	0.09	0.06	n/a	0.10	0.11	0.05	0.24	0.05	0.09
AUT	0.54	0.06	0.10	0.07	0.10	0.11	0.05	n/a	0.16	0.14	0.05	0.33	0.02	0.20
BEL	0.55	0.06	0.04	0.05	0.05	0.11	0.07	0.08	0.12	0.11	0.02	0.29	0.03	0.19
CAN	0.20	0.04	0.05	0.04	0.04	0.08	0.05	n/a	0.08	0.08	n/a	0.16	0.04	0.07
CHE	0.34	0.06	0.06	0.06	0.09	0.09	0.04	n/a	0.09	0.10	0.04	0.17	0.06	0.13
CHL	0.55	0.04	0.11	0.03	0.04	0.18	0.08	n/a	0.13	0.15	0.12	0.64	0.04	0.09
CZE	0.56	0.06	0.07	0.06	0.05	0.24	0.11	n/a	0.12	0.26	0.07	0.27	0.11	0.19
DEU	0.55	0.03	0.04	0.04	0.05	0.05	0.03	n/a	0.10	0.07	0.03	0.27	0.01	0.16
DNK	0.18	0.04	0.08	0.03	0.03	0.14	0.08	0.45	0.09	0.13	0.03	0.18	0.02	0.03
ESP	0.17	0.03	0.05	0.05	0.06	0.10	0.07	n/a	0.09	0.11	0.03	0.22	0.05	0.06
EST	0.19	0.03	0.08	0.03	0.06	0.21	0.11	n/a	0.08	0.17	0.02	0.23	0.08	0.06
FIN	0.13	0.02	0.04	0.03	0.05	0.07	0.05	0.41	0.04	0.08	0.02	0.12	0.03	0.04
FRA	0.58	0.03	0.03	0.03	0.04	0.10	0.04	n/a	0.20	0.11	n/a	0.28	0.03	0.13
GBR	0.24	0.05	0.06	0.05	0.05	0.08	0.05	n/a	0.08	0.08	n/a	0.17	0.05	0.06
GRC	0.39	0.05	0.05	0.02	0.02	0.12	0.04	n/a	0.13	0.12	0.02	0.31	0.03	0.09
HUN	0.65	0.10	0.10	0.05	0.06	0.15	0.10	n/a	0.28	0.24	0.06	0.40	0.03	0.19
IRL	0.19	0.03	0.03	0.01	0.02	0.08	0.02	n/a	0.06	0.10	0.03	0.21	0.03	0.05
ISL	0.13	0.02	0.07	0.04	0.04	0.20	0.07	n/a	0.07	0.18	0.00	0.15	0.04	0.01
ISR	0.43	0.03	0.06	0.08	0.04	0.05	0.03	n/a	0.11	0.08	0.06	0.27	0.04	0.08
ITA	0.53	0.05	0.07	0.05	0.06	0.11	0.10	0.23	0.15	0.16	0.05	0.24	0.05	0.17
JPN	0.54	0.03	0.04	0.04	0.04	0.09	0.10	n/a	0.13	0.19	0.02	0.23	0.07	0.16
KOR	0.39	0.02	0.09	0.07	0.06	0.10	0.09	0.03	0.10	0.17	0.02	0.22	0.06	0.20
LUX	0.33	0.03	0.03	0.02	0.06	0.03	0.02	n/a	0.13	0.04	n/a	0.28	0.00	0.04
MEX	0.36	0.04	0.07	0.05	0.06	0.10	0.08	n/a	0.14	0.11	0.28	0.47	0.04	0.07
NLD	0.67	0.03	0.05	0.03	0.05	0.10	0.07	n/a	0.08	0.09	0.01	0.18	0.02	0.32
NOR	0.15	0.02	0.04	0.01	0.03	0.09	0.06	n/a	0.07	0.13	0.00	0.09	0.05	0.05
NZL	0.25	0.06	0.04	0.02	0.01	0.05	0.04	n/a	0.07	0.09	0.04	0.24	0.02	0.06
POL	0.25	0.04	0.06	0.03	0.05	0.17	0.07	n/a	0.08	0.18	0.02	0.28	0.03	0.05
PRT	0.32	0.02	0.05	0.05	0.04	0.12	0.05	0.06	0.13	0.09	0.01	0.30	0.02	0.05
SVK	0.46	0.07	0.07	0.06	0.07	0.18	0.10	0.02	0.13	0.17	0.12	0.37	0.07	0.10
SVN	0.59	0.05	0.14	0.09	0.08	0.16	0.07	0.09	0.14	0.22	0.05	0.25	0.05	0.14
SWE	0.15	0.02	0.05	0.02	0.03	0.10	0.07	n/a	0.08	0.11	0.01	0.14	0.03	0.06
TUR	0.62	0.04	0.06	0.03	0.05	0.10	0.04	n/a	0.11	0.10	0.11	0.30	0.02	0.11
USA	0.25	0.02	0.04	0.02	0.02	0.05	0.02	n/a	0.06	0.08	n/a	0.28	0.01	0.04

Note: n/a indicates country not administering questions forming construct.

[Part 2/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	FAILMAT	FAMCON	FAMCONC	HEDRES	HERITCUL	HISEI	HOMEPOS	HOMSCH	HOSTCUL	ICTATTNEG	ICTATTPOS	ICTHOME	ICTSCH
AUS	0.05	0.14	0.13	0.09	n/a	0.14	0.17	0.20	n/a	0.04	0.06	0.03	0.14
AUT	0.02	0.32	0.32	0.12	n/a	0.27	0.17	0.23	n/a	0.06	0.08	0.05	0.14
BEL	0.10	0.21	0.23	0.17	0.07	0.24	0.19	0.13	0.09	0.02	0.06	0.02	0.24
CAN	0.05	0.07	0.10	0.07	n/a	0.10	0.13	n/a	n/a	n/a	n/a	n/a	n/a
CHE	0.05	0.25	0.19	0.05	n/a	0.14	0.10	0.11	n/a	0.03	0.05	0.03	0.27
CHL	0.04	0.29	0.21	0.18	n/a	0.52	0.56	0.11	n/a	0.04	0.03	0.26	0.11
CZE	0.06	0.23	0.19	0.13	n/a	0.24	0.14	0.18	n/a	0.08	0.05	0.04	0.14
DEU	0.04	0.21	0.29	0.11	n/a	0.26	0.12	0.11	n/a	0.03	0.03	0.03	0.20
DNK	0.04	0.09	0.06	0.09	0.26	0.12	0.13	0.14	0.41	0.04	0.03	0.03	0.09
ESP	0.04	0.09	0.15	0.04	n/a	0.18	0.12	0.09	n/a	0.09	0.03	0.04	0.15
EST	0.08	0.10	0.08	0.06	n/a	0.18	0.12	0.11	n/a	0.06	0.04	0.04	0.11
FIN	0.03	0.12	0.05	0.05	0.56	0.09	0.06	0.08	0.44	0.02	0.03	0.02	0.14
FRA	0.02	0.19	0.24	0.07	n/a	0.22	0.19	n/a	n/a	n/a	n/a	n/a	n/a
GBR	0.03	0.12	0.09	0.06	n/a	0.12	0.12	n/a	n/a	n/a	n/a	n/a	n/a
GRC	0.02	0.08	0.12	0.10	n/a	0.24	0.21	0.06	n/a	0.03	0.02	0.07	0.09
HUN	0.05	0.23	0.31	0.19	n/a	0.30	0.31	0.04	n/a	0.11	0.02	0.09	0.06
IRL	0.02	0.06	0.06	0.05	n/a	0.13	0.15	0.06	n/a	0.01	0.02	0.03	0.11
ISL	0.02	0.05	0.06	0.05	n/a	0.13	0.07	0.09	n/a	0.01	0.05	0.02	0.19
ISR	0.01	0.16	0.11	0.09	n/a	0.25	0.21	0.19	n/a	0.09	0.08	0.08	0.15
ITA	0.04	0.25	0.25	0.07	0.18	0.21	0.16	0.06	0.20	0.06	0.04	0.04	0.16
JPN	0.02	0.25	0.19	0.11	n/a	0.12	0.15	0.08	n/a	0.01	0.03	0.07	0.13
KOR	0.01	0.35	0.24	0.11	0.16	0.12	0.15	0.11	0.09	0.01	0.05	0.07	0.07
LUX	0.01	0.10	0.11	0.04	n/a	0.31	0.11	n/a	n/a	n/a	n/a	n/a	n/a
MEX	0.05	0.17	0.08	0.21	n/a	0.31	0.43	0.33	n/a	0.04	0.07	0.33	0.22
NLD	0.02	0.20	0.19	0.08	n/a	0.15	0.10	0.11	n/a	0.06	0.02	0.02	0.13
NOR	0.02	n/a	n/a	0.04	n/a	0.07	0.07	0.15	n/a	0.01	0.03	0.02	0.14
NZL	0.02	0.14	0.08	0.09	n/a	0.17	0.19	0.13	n/a	0.05	0.02	0.02	0.13
POL	0.05	0.08	0.07	0.02	n/a	0.25	0.14	0.05	n/a	0.03	0.01	0.04	0.13
PRT	0.04	0.08	0.10	0.08	0.09	0.27	0.17	0.03	0.04	0.06	0.01	0.03	0.10
SVK	0.03	0.19	0.17	0.38	0.07	0.25	0.37	0.17	0.12	0.07	0.03	0.13	0.14
SVN	0.03	0.19	0.21	0.07	0.11	0.22	0.14	0.07	0.14	0.05	0.04	0.06	0.10
SWE	0.02	0.05	0.06	0.05	n/a	0.12	0.08	0.15	n/a	0.02	0.03	0.02	0.17
TUR	0.03	0.15	0.22	0.25	n/a	0.16	0.35	0.05	n/a	0.07	0.02	0.26	0.09
USA	0.02	0.09	0.12	0.10	n/a	0.17	0.19	n/a	n/a	n/a	n/a	n/a	n/a

Note: n/a indicates country not administering questions forming construct.

[Part 3/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	INFOCAR	INFOJOB1	INFOJOB2	INSTMOT	INTMAT	MATBEH	MATHEFF	MATINTFC	MATWKETH	MTSUP	OPENPS	OUTHOURS	PARED	PERSEV
AUS	0.10	0.12	0.08	0.03	0.05	0.07	0.13	0.04	0.06	0.07	0.05	0.08	0.14	0.04
AUT	0.12	0.04	0.00	0.13	0.07	0.05	0.19	0.14	0.07	0.07	0.04	0.07	0.22	0.03
BEL	0.06	0.04	0.03	0.06	0.05	0.06	0.12	0.06	0.04	0.07	0.04	0.07	0.16	0.04
CAN	0.08	0.17	0.10	0.04	0.04	0.07	0.08	0.04	0.05	0.06	0.04	0.09	0.09	0.05
CHE	n/a	n/a	n/a	0.07	0.04	0.04	0.13	0.07	0.05	0.06	0.03	0.05	0.10	0.04
CHL	n/a	n/a	n/a	0.03	0.04	0.03	0.07	0.04	0.04	0.10	0.03	0.05	0.45	0.03
CZE	n/a	n/a	n/a	0.09	0.09	0.10	0.16	0.12	0.07	0.18	0.07	0.07	0.16	0.04
DEU	n/a	n/a	n/a	0.01	0.03	0.03	0.12	0.03	0.01	0.09	0.02	0.05	0.17	0.01
DNK	0.11	0.06	0.03	0.03	0.04	0.04	0.08	0.05	0.05	0.08	0.06	0.06	0.10	0.04
ESP	n/a	n/a	n/a	0.03	0.04	0.04	0.06	0.04	0.04	0.06	0.03	0.03	0.16	0.05
EST	n/a	n/a	n/a	0.02	0.06	0.06	0.07	0.03	0.07	0.12	0.02	0.08	0.14	0.03
FIN	0.05	0.07	0.08	0.04	0.04	0.02	0.04	0.02	0.03	0.05	0.02	0.03	0.06	0.03
FRA	n/a	n/a	n/a	0.02	0.05	0.02	0.15	0.02	0.04	0.09	0.05	0.07	0.14	0.03
GBR	n/a	n/a	n/a	0.04	0.05	0.07	0.09	0.03	0.04	0.06	0.04	0.10	0.08	0.04
GRC	n/a	n/a	n/a	0.03	0.06	0.04	0.08	0.04	0.06	0.08	0.04	0.09	0.20	0.05
HUN	0.12	0.07	0.03	0.08	0.08	0.07	0.26	0.08	0.05	0.12	0.09	0.09	0.28	0.04
IRL	0.06	0.09	0.05	0.04	0.03	0.02	0.07	0.04	0.04	0.04	0.00	0.07	0.11	0.01
ISL	n/a	n/a	n/a	0.01	0.03	0.02	0.05	0.02	0.04	0.11	0.02	0.04	0.08	0.01
ISR	n/a	n/a	n/a	0.07	0.10	0.16	0.09	0.05	0.06	0.06	0.04	0.05	0.17	0.06
ITA	0.05	0.05	0.03	0.09	0.11	0.09	0.15	0.11	0.09	0.07	0.06	0.13	0.14	0.06
JPN	n/a	n/a	n/a	0.09	0.08	0.08	0.25	0.03	0.05	0.08	0.07	0.20	0.22	0.05
KOR	0.10	0.11	0.10	0.11	0.09	0.15	0.22	0.05	0.11	0.08	0.08	0.19	0.15	0.04
LUX	0.11	0.10	0.01	0.03	0.03	0.03	0.09	0.02	0.01	0.05	0.03	0.01	0.17	0.01
MEX	n/a	n/a	n/a	0.05	0.08	0.06	0.07	0.05	0.06	0.08	0.06	0.06	0.34	0.03
NLD	n/a	n/a	n/a	0.04	0.04	0.04	0.11	0.03	0.02	0.07	0.02	0.06	0.08	0.02
NOR	n/a	n/a	n/a	0.01	0.03	0.02	0.04	0.04	0.02	0.05	0.01	0.03	0.06	0.02
NZL	0.09	0.08	0.01	0.02	0.06	0.09	0.11	0.02	0.03	0.03	0.03	0.08	0.11	0.03
POL	n/a	n/a	n/a	0.04	0.04	0.07	0.09	0.02	0.05	0.11	0.02	0.03	0.21	0.03
PRT	0.06	0.06	0.07	0.05	0.05	0.03	0.14	0.03	0.05	0.05	0.04	0.04	0.22	0.04
SVK	0.10	0.04	0.03	0.07	0.10	0.09	0.16	0.09	0.06	0.09	0.04	0.07	0.22	0.05
SVN	0.07	0.08	0.09	0.10	0.08	0.06	0.15	0.13	0.05	0.08	0.04	0.26	0.17	0.03
SWE	n/a	n/a	n/a	0.03	0.02	0.02	0.07	0.02	0.03	0.05	0.04	0.05	0.06	0.01
TUR	n/a	n/a	n/a	0.04	0.05	0.07	0.18	0.02	0.05	0.04	0.04	0.05	0.19	0.02
USA	n/a	n/a	n/a	0.01	0.05	0.04	0.08	0.02	0.03	0.02	0.02	0.07	0.22	0.02

Note: n/a indicates country not administering questions forming construct.

[Part 4/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	SCMAT	SMINS	STUDREL	SUBNORM	TCHBEHFA	TCHBEHSO	TCHBEHTD	TEACHSUP	USEMATH	USESCH	WEALTH	AVG*
AUS	0.05	0.16	0.09	0.07	0.06	0.11	0.05	0.06	0.08	0.20	0.13	0.09
AUT	0.05	0.31	0.08	0.07	0.10	0.18	0.05	0.12	0.09	0.16	0.09	0.13
BEL	0.04	0.20	0.03	0.06	0.05	0.14	0.08	0.08	0.16	0.22	0.10	0.11
CAN	0.05	0.23	0.06	0.06	0.06	0.10	0.06	0.06	n/a	n/a	0.13	0.08
CHE	0.03	0.14	0.09	0.08	0.08	0.24	0.06	0.07	0.09	0.13	0.07	0.09
CHL	0.03	0.11	0.06	0.03	0.13	0.17	0.11	0.12	0.15	0.14	0.58	0.15
CZE	0.07	0.20	0.09	0.11	0.10	0.14	0.11	0.18	0.14	0.14	0.09	0.13
DEU	0.01	0.28	0.08	0.04	0.09	0.15	0.05	0.09	0.07	0.16	0.08	0.10
DNK	0.03	0.09	0.10	0.07	0.06	0.09	0.07	0.11	0.22	0.09	0.12	0.09
ESP	0.04	0.08	0.07	0.04	0.08	0.10	0.07	0.10	0.10	0.16	0.10	0.08
EST	0.04	0.11	0.08	0.03	0.10	0.18	0.16	0.12	0.10	0.12	0.08	0.09
FIN	0.02	0.11	0.05	0.06	0.05	0.08	0.06	0.07	0.10	0.10	0.05	0.08
FRA	0.03	0.12	0.04	0.04	0.08	0.12	0.08	0.10	n/a	n/a	0.10	0.10
GBR	0.05	0.18	0.06	0.06	0.06	0.12	0.07	0.06	n/a	n/a	0.09	0.08
GRC	0.04	0.13	0.07	0.06	0.09	0.17	0.06	0.10	0.09	0.09	0.15	0.09
HUN	0.07	0.40	0.07	0.13	0.15	0.25	0.13	0.18	0.14	0.12	0.15	0.15
IRL	0.02	0.11	0.04	0.03	0.06	0.10	0.05	0.05	0.08	0.11	0.10	0.06
ISL	0.02	0.17	0.07	0.05	0.04	0.09	0.12	0.13	0.12	0.19	0.06	0.07
ISR	0.02	0.09	0.09	0.09	0.15	0.15	0.06	0.07	0.20	0.12	0.30	0.11
ITA	0.08	0.29	0.08	0.08	0.11	0.15	0.10	0.12	0.15	0.13	0.10	0.12
JPN	0.03	0.60	0.06	0.10	0.05	0.11	0.08	0.08	0.02	0.12	0.10	0.12
KOR	0.07	0.73	0.08	0.14	0.05	0.06	0.05	0.05	0.05	0.17	0.08	0.12
LUX	0.01	0.27	0.03	0.03	0.05	0.11	0.03	0.06	n/a	n/a	0.05	0.07
MEX	0.04	0.13	0.06	0.06	0.11	0.13	0.10	0.12	0.08	0.16	0.45	0.14
NLD	0.03	0.04	0.04	0.04	0.08	0.16	0.07	0.07	0.10	0.11	0.05	0.09
NOR	0.02	0.11	0.05	0.04	0.07	0.08	0.07	0.08	0.12	0.14	0.05	0.06
NZL	0.02	0.16	0.06	0.07	0.05	0.10	0.05	0.04	0.08	0.13	0.14	0.08
POL	0.03	0.85	0.09	0.05	0.08	0.11	0.10	0.14	0.14	0.11	0.13	0.10
PRT	0.03	0.27	0.05	0.06	0.09	0.14	0.08	0.09	0.09	0.12	0.11	0.09
SVK	0.07	0.44	0.10	0.14	0.14	0.22	0.11	0.15	0.18	0.13	0.21	0.14
SVN	0.06	0.34	0.08	0.08	0.12	0.21	0.10	0.13	0.10	0.12	0.07	0.12
SWE	0.02	0.39	0.06	0.06	0.04	0.07	0.04	0.07	0.08	0.23	0.07	0.07
TUR	0.04	0.24	0.03	0.05	0.04	0.13	0.04	0.04	0.14	0.13	0.29	0.11
USA	0.03	0.11	0.06	0.03	0.05	0.12	0.04	0.03	n/a	n/a	0.19	0.07

*Average rho across all questionnaire scales in this country.
Note: n/a indicates country not administering questions forming construct.

Still, consideration has to be given to the possibilities of analysing responses to questions in the rotated part by subgroups of interest, such as immigrants or single-parents. Such analyses will have become more limited as group sizes decreased. In general, in PISA results are not reported if there are fewer than 30 students or fewer than five schools with valid data.

To explore the implications for analyses on correlations and corresponding standard errors, correlations between questionnaire indices and mathematics achievement computed using all student responses were compared to the correlation computed using 2/3 of the student responses. PISA 2003 data for Finland, Germany and Korea were used as these countries differ in terms of their intraclass correlation and geolocation. Thus, Germany has a high intraclass correlation with values differing across schools but tending to be similar for students in the same school whereas the reverse applies in Finland. Results of the analyses are shown in Table 17.11.

Results in Table 17.11 confirm earlier analyses (Adams, Lietz and Berezner, 2013; Berezner and Lietz, 2009) whereby the differences in correlation coefficients range from 0 to 0.02 and the corresponding standard errors differ from 0 to 0.01. Therefore, differences are considered negligible.

It should be noted that these comparative analyses could not be replicated using PISA 2012 data as, by design, information was available for only two thirds of students per question in the rotated parts of the student questionnaire in that cycle.

One alternative to assigning the missing data due to rotation a specific code (i.e. ‘7’) would have been to attempt imputations as a means of replacing missing information with ‘pseudo-information’. However, a number of considerations with respect to the optimal model, challenges from bidirectional imputation, logical inconsistencies in data patterns and precision led to the decision not to impute the values that were missing due to rotation in PISA 2012.

Table 17.11 Correlations between questionnaire indices and mathematics achievement (PISA 2003 data)

		All students		2/3 students		
CNT	Index	Corr	SE	Corr	SE	Increase in SE (ratio: Col6/Col4)
DEU	ANXMAT	-0.34	0.015	-0.35	0.018	1.20
DEU	ATSCHL	-0.10	0.019	-0.09	0.022	1.15
DEU	ATTCOMP	0.00	0.015	-0.02	0.022	1.49
DEU	BELONG	-0.02	0.018	-0.02	0.020	1.09
DEU	COMPHOME	0.26	0.016	0.25	0.022	1.39
DEU	COMPLRN	-0.06	0.018	-0.06	0.021	1.19
DEU	COOPLRN	-0.03	0.016	-0.04	0.019	1.14
DEU	CSTRAT	-0.08	0.020	-0.09	0.024	1.18
DEU	CULTPOS	0.25	0.014	0.27	0.018	1.27
DEU	DISCLIM	0.22	0.019	0.24	0.022	1.14
DEU	ELAB	-0.06	0.018	-0.05	0.022	1.24
DEU	ESCS	0.48	0.014	0.49	0.016	1.17
DEU	HEDRES	0.25	0.015	0.26	0.019	1.26
DEU	HIGHCONF	0.09	0.021	0.09	0.026	1.25
DEU	HOMEPOS	0.39	0.012	0.40	0.017	1.43
DEU	INSTMOT	0.02	0.019	0.03	0.023	1.20
DEU	INTCONF	0.21	0.017	0.21	0.020	1.22
DEU	INTMAT	0.12	0.019	0.14	0.020	1.10
DEU	INTUSE	0.05	0.019	0.03	0.025	1.33
DEU	MATHEFF	0.50	0.015	0.51	0.017	1.19
DEU	MEMOR	-0.22	0.019	-0.22	0.025	1.35
DEU	PRGUSE	-0.06	0.022	-0.06	0.027	1.25
DEU	ROUTCONF	0.32	0.015	0.32	0.016	1.04
DEU	SCMAT	0.27	0.016	0.29	0.018	1.14
DEU	STUREL	-0.07	0.021	-0.05	0.023	1.09
DEU	TEACHSUP	-0.12	0.020	-0.11	0.023	1.15
FIN	ANXMAT	-0.44	0.013	-0.44	0.015	1.16
FIN	ATSCHL	0.14	0.015	0.14	0.017	1.12
FIN	ATTCOMP	-0.01	0.013	-0.01	0.016	1.20
FIN	BELONG	-0.03	0.015	-0.03	0.018	1.21
FIN	COMPHOME	0.15	0.015	0.16	0.019	1.26
FIN	COMPLRN	0.23	0.015	0.24	0.019	1.29
FIN	COOPLRN	0.03	0.017	0.05	0.020	1.21
FIN	CSTRAT	0.11	0.013	0.13	0.019	1.53
FIN	CULTPOS	0.21	0.015	0.21	0.016	1.07
FIN	DISCLIM	0.12	0.015	0.12	0.020	1.33
FIN	ELAB	0.17	0.014	0.17	0.018	1.27
FIN	ESCS	0.33	0.014	0.32	0.016	1.13
FIN	HEDRES	0.18	0.017	0.18	0.020	1.20
FIN	HIGHCONF	0.09	0.015	0.09	0.019	1.27
FIN	HOMEPOS	0.27	0.015	0.27	0.018	1.15
FIN	INSTMOT	0.29	0.017	0.30	0.019	1.13
FIN	INTCONF	0.11	0.017	0.12	0.018	1.10
FIN	INTMAT	0.33	0.015	0.34	0.019	1.25
FIN	INTUSE	0.04	0.013	0.05	0.018	1.32
FIN	MATHEFF	0.52	0.013	0.51	0.015	1.21
FIN	MEMOR	0.08	0.017	0.08	0.023	1.38
FIN	PRGUSE	0.03	0.015	0.04	0.019	1.22
FIN	ROUTCONF	0.23	0.014	0.23	0.018	1.31
FIN	SCMAT	0.57	0.011	0.58	0.014	1.25
FIN	STUREL	0.10	0.014	0.11	0.018	1.26
FIN	TEACHSUP	0.05	0.018	0.04	0.021	1.18
KOR	ANXMAT	-0.22	0.014	-0.23	0.017	1.20
KOR	ATSCHL	0.00	0.016	0.01	0.016	0.99
KOR	ATTCOMP	0.02	0.018	0.03	0.021	1.13
KOR	BELONG	0.10	0.017	0.09	0.019	1.07
KOR	COMPHOME	0.26	0.015	0.26	0.019	1.26
KOR	COMPLRN	0.40	0.013	0.40	0.018	1.32
KOR	COOPLRN	0.34	0.014	0.34	0.016	1.16
KOR	CSTRAT	0.40	0.014	0.39	0.016	1.12
KOR	CULTPOS	0.27	0.022	0.26	0.025	1.14
KOR	DISCLIM	0.13	0.017	0.14	0.021	1.19
KOR	ELAB	0.30	0.014	0.31	0.018	1.29
KOR	ESCS	0.37	0.025	0.38	0.026	1.05
KOR	HEDRES	0.26	0.018	0.26	0.021	1.17
KOR	HIGHCONF	0.11	0.016	0.13	0.017	1.05
KOR	HOMEPOS	0.36	0.021	0.36	0.025	1.17
KOR	INSTMOT	0.35	0.015	0.35	0.016	1.10
KOR	INTCONF	0.20	0.016	0.19	0.021	1.30
KOR	INTMAT	0.39	0.014	0.40	0.016	1.13
KOR	INTUSE	-0.07	0.018	-0.08	0.021	1.16
KOR	MATHEFF	0.58	0.012	0.58	0.016	1.30
KOR	MEMOR	0.18	0.016	0.19	0.019	1.13
KOR	PRGUSE	0.08	0.016	0.10	0.021	1.30
KOR	ROUTCONF	0.32	0.014	0.33	0.016	1.11
KOR	SCMAT	0.46	0.014	0.47	0.016	1.16
KOR	STUREL	0.10	0.024	0.12	0.022	0.95
KOR	TEACHSUP	0.06	0.019	0.05	0.019	1.00

Imputations for missing data are model-dependent draws from the posterior distribution of random variables, conditional on the observed values of other available variables, using estimated relationships between the variable that is missing and the remainder of the variables. Many possible imputation models are conceivable which differ, among other things, in terms of the distributional assumptions that are made, the number of dimensions and the extent to which information from other observed student and school-level variables is used.

To do imputations properly for the whole of the international database would require an IRT (Item Response Theory) model with the number of dimensions equal to the reported proficiency scales plus the number of questionnaire scales with another dimension added for each missing background variable. The scope of these imputations would be well beyond what has been done to date.

The problem of bidirectional imputation would arise if the imputation for missing background data were to include proficiency data. Here, a problem of circularity may arise as plausible values for proficiencies are imputed using information from the background variables in the first place.

In the absence of imputed values for the student questionnaire data that are missing due to rotation, a few alternatives are recommended when using data from all three question sets. Central to the justification for these alternatives is that the missing data that results from rotation is missing completely at random.

First, since the design produces an estimate of the complete variance/covariance matrix modelling that utilises only the variance/covariance matrix can easily be undertaken.

Second, where possible full information maximum likelihood (FIML) can be implemented. Good introductions to FIML have been written by Marcoulides and Schumacker (1996) and Wothke (2000), among others. Software programs that allow FIML estimation include Amos and MPlus.

Third, analysts can impute data required for the specific models they seek to analyse. Software programs that enable imputation include MPlus, R, SAS®, SPSS® and Stata.

Intentions of the analyses

If the intended analyses of the PISA 2012 data are multilevel in nature, consideration needs to be given to any variables and constructs for which data were collected at the student level with the intention of using them as contextual variables at the school level. Thus, for example, individual students' ratings of their school's disciplinary climate are frequently aggregated and entered at the school level in multilevel models examining differences in student achievement. Lüdtke et al. (2009) have pointed out that the reliability of the observed group mean rating – labelled ICC(2) – is dependent on the proportion of the total variance that can be attributed to between-group differences – labelled ICC(1) – and the number of students from whom ratings have been obtained. The following equations define ICC(1) and ICC(2):

17.1

$$ICC(1) = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W}$$

17.2

$$ICC(2) = \frac{k \times ICC(1)}{1 + (k-1) \times ICC(1)}$$

Where k is the number of students in each group; MS_B is the between mean square; and MS_w is the within mean square from analysis of variance.

In their article, Lüdtke et al. (2009) assumed a rather high ICC(1) of 0.43 indicating that 43% of the total variation found in all student ratings was associated with differences between groups. Combined with a group size of ten students, this translated to a reliability of the group mean rating ICC(2) of 0.88 which was a relatively high level of reliability for an attitudinal scale.

In Table 17.12, implications for the reliability of the group mean rating ICC(2) is given for a number of different combinations of ICC(1)s and numbers of students.

Table 17.12 Implications of different group sizes on reliabilities of group mean ratings

ICC(1)	ICC(2) for 20 students (= 3 rotated forms)	ICC(2) for 30 students = no rotation
0.43	0.94	0.96
0.30	0.90	0.93
0.25	0.87	0.91
0.20	0.83	0.88
0.15	0.78	0.84
0.10	0.69	0.77
0.05	0.51	0.61

Note: For definitions of ICC(1) and ICC(2) see equations 1 and 2 above.

Table 17.12 demonstrates that the rotation design has had different implications depending on the between-group differences associated with a particular scale or construct – which is likely to differ for the countries participating in PISA.

In such contextual analyses, Lüdtke et al. (2008) have recommended to use a multilevel latent covariate approach (MLC) instead of a multilevel manifest covariate (MMC) approach, particularly when a) the ICC(1) is small, b) the sampling proportion (i.e. in the PISA context the number of 15-year old students that the 30 randomly selected 15-year-olds in each school represent) is small and c) the number of cases are small. Through a number of simulation and real-data applications, the authors showed that the MLC could correct for the unreliability at level 2 for constructs that represented aggregates of student-level ratings. The authors particularly argued for the use of the MLC approach where the L2 construct was assumed to be a reflective aggregation of L1 measures. Reflective constructs, for example, would be individual students' ratings of instructional practices of the teacher (a common referent at L2). Formative L2 constructs for which no such adjustment might be necessary would include aggregations of L1 variables such as age, gender, or SES (socio-economic status).

In line with Lüdtke et al. (2009), it is recommended that future multilevel analyses of student-level ratings of instructional practices or school context at the between-school level should report both ICC(1) and ICC(2).

Analyses were conducted with data for five countries from PISA 2003, the cycle prior to 2012 in which mathematics had been the major domain (Van de gaer, Lietz and Adams, 2011). The countries, namely Australia, Finland, Germany, Korea and the United States were selected as they differed with respect to the following characteristics which Lüdtke et al. (2008) had specified as impacting on the group level effects:

- The number of level 1 cases per group (from 19 in the United States to 38 in Australia);
- The number of groups (from 149 in Korea to 321 in Australia);
- The ICC in the outcome variable, mathematics achievement (ranging from 0.06 in Finland to 0.63 in Germany).

In addition, these five countries differed in terms of their mathematics achievement in 2003 relative to the OECD average, with Finland, Korea and Australia performing above average, Germany around average and the United States below average. Finally, these countries also included some cultural variation.

The purpose of the analyses was to examine the differences between the multilevel latent covariate approach (MLC) and the multilevel manifest covariate approach (MMC) to modelling individual students' ratings of instructional practices or school context at the between-school level, particularly for reflective constructs where individual students' ratings related to a common referent at level 2. At the same time, the analyses explored the implications of the rotated student questionnaire design which resulted in having only two thirds of student responses available in the analyses. Results showed that the differences in estimates between the unrotated and the rotated questionnaire design were far smaller than the differences in estimates between the multilevel latent covariate approach (MLC) and the multilevel manifest covariate approach (MMC).

In addition, the results of the analyses seemed to confirm partially the findings of Lüdtke et al. (2008). Thus, the larger beta coefficients of the group-level effects using the multilevel latent covariate approach seemed to suggest that the unreliability of the group-level effects was taken into account. The fact that the SEs (standard errors) of the MMC approach in most cases were smaller than those of the MLC approach might have been expected because the group mean of the covariate was treated as observed whereas in the MLC it was not (leading to larger sampling variability).

Results showed the most pronounced difference in the beta coefficient of the group-level variable in the United States between the MLC and MMC approach. This could be expected as, according to Lüdtke et al. (2008), the difference

between the two approaches (MMC and MLC) were the most pronounced for small numbers of level 1 cases within each level 2 group and the United States had the lowest cluster size (=19) of the countries included in the analysis.

In other words, the MLC approach seemed to control for the unreliability in the group-level effects that was introduced when a reduced number of students were being sampled within schools. Using the MMC approach seemed to result in an underestimation of the group-level effect. Thus, the approaches MLC and MMC seemed to lead to different estimates of the group-level effect in multilevel analyses and more so for countries with low predictor ICC, relatively higher outcome ICC and relatively smaller average cluster size.

While differences were noted between the absolute sizes of the estimates for rotated and unrotated questionnaire designs as well as and the analytical approaches (i.e. MMC compared with MLC), the conclusions in terms of whether or not an effect would be considered significant remained the same in all instances.

Still, as Lüdtke et al. (2008) noted, most analysts use the MMC approach. Where analysts of PISA data have used this approach previously and want to replicate their analyses with PISA 2012 data, they should also use the MMC approach with the PISA 2012 data to ensure comparability of results. Alternatively, they may choose to rerun the earlier analysis using the MLC approach and also apply this approach to the PISA 2012 data.

Notes

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

2. This link is not available anymore, but for more details on the correspondance between ISCO-88 and ISCO-08 please refer to the following website: *http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm*.

Reference

Adams, R. J., P. Lietz and **A. Berezner** (2013), "On the use of rotated context questionnaires in conjunction with multilevel item response models", *Large-scale Assessments in Education*, No. 1(1).

Bartram, D. (2007), "Increasing validity with forced-choice criterion measurement formats", *International Journal of Selection and Assessment*, No. 15, pp. 263-272. *http://dx.doi.org/10.1111/j.1468-2389.2007.00386.x*

Bertling, J. P. and **P.C. Kyllonen** (2013), "Anchoring adjustments of student questionnaire indexes: possible scenarios for the PISA 2012 international database", Paper prepared for the PISA Technical Advisory Group (TAG) meeting, Melbourne, 4-5 April 2013.

Bertling, J. and **R. Roberts** (2011), "Student questionnaire (Opportunity to learn): Field Trial analyses", Paper presented at the QEG Meeting, Melbourne (QEG(1109)1b), Australia, 26-28 September 2011.

Berezner, A. and **P. Lietz** (2009), "Implication for Preferred Model of Rotation on Cognitive Proficiency Estimate Generation Reporting and Subsequent Analyses", Paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October 2009.

Bowen, C.C., B.A. Martin and **S.T. Hunt** (2002), "A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires", International *Journal of Organizational Analysis*, No. 10(3), pp. 240-259.

Brown, A. and **A. Maydeu-Olivares** (2013), "How IRT can solve problems of ipsative data in forced-choice questionnaires", *Psychological Methods*, No. 18(1), p. 36.

Cheung, M.W.L. and **W. Chan** (2002), “Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis”, *Structural Equation Modeling,* No. 9, pp. 55-77.

Christiansen, N., G. Burns and **G. Montgomery** (2005), “Reconsidering the use of forced-choice formats for applicant personality assessment”, *Human Performance*, No. 18, pp. 267–307. *http://dx.doi.org/10.1207/s15327043hup1803_4*

Christian, M. S., B. D. Edwards and **J. C. Bradley**, (2010), “Situational judgment tests: Constructs assessed and a meta-analytic of their criterion-related validities”, *Personnel Psychology*, No. 63, pp. 83-117.

De la Torre, J., V.Ponsoda, I. Leenen and **P. Hontangas** (2011), “Some extensions of the Multi-Unidimensional Pairwise Preference Model”*,* Paper presented at the 26th annual conference for the Society of Industrial and Organizational Psychology, Chicago.

Fischer, R., J. Fontaine, F. J. R. van de Vijver and **D. van Hemert** (2009), “What is style and what is bias in cross-cultural comparisons? An examination of response styles in cross-cultural research”, in A. Gari and K. Mylonas (eds.), *Quod Erat Demonstrandum: From Herodotus’ Ethnographic Journeys to Cross-Cultural Research*, Pedio, Athens, pp. 137 – 148.

Ganzeboom, Harry B.G. (2010), “Do’s and Don’t’s of Occupation Coding with an Extension on ISCO-08”, Working paper, Department of Social Research Methodology, VU Amsterdam, last revised January 3 2010.

Hopkins, D. J. and **G. King** (2010), “Improving anchoring vignettes designing surveys to correct interpersonal incomparability, *Public Opinion Quarterly*.

Kyllonen, P., J. Bertling and **R. Roberts** (2013), “Overclaiming: A new technique for correcting for response tendencies in international assessments”, Paper presented at the 15th Biennial Conference EARLI 2013, Munich, Germany.

Kyllonen, P., P. Lietz and **R. Roberts** (2010), “Cross-cultural comparative questionnaire issues », Paper prepared for the PISA2012 Questionnaire Expert Group (QEG) meeting, Hong Kong, 25-27 February 2010.

Lievens, F., T. Buyse and **P. R. Sackett** (2005), “The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains”, *Journal of Applied Psychology*, No. 90, pp. 442-452.

Lievens, F. and **P. Coestsier** (2002), “Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity”, *International Journal of Selection and Assessment*, No.10, pp. 245-257.

Lüdtke, O., H. W. Marsh, A. Robitzsch, U.Trautwein, T. Asparouhov and **B. Muthén** (2008), “The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies”, *Psychological Methods,* No. 13, pp. 203-229.

Lüdtke, O., A. Robitzsch, U. Trautwein and **M. Kunter** (2009), “Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modelling”, *Contemporary Educational Psychology*, No. 34(2), pp. 120-131.

Maydeu-Olivares, A. and **A. Brown** (2010), “Item response modelling of paired comparison and ranking data”, *Multivariate Behavioural Research,* No. 45, pp. 935-974.
http://dx.doi.org 10.1080/00273171.2010.531231

McDaniel, M.A., F.P. Morgeson, E.B. Finnegan, M.A. Campion and **E.P. Braverman** (2001), “Predicting job performance using situational judgment tests: A clarification of the literature”, *Journal of Applied Psychology*, No. 86, pp. 730-740.

McDaniel, M. A., J. L Anderson., M. H. Derbish and **N. Morrisette** (2007), “Testing the testing effect in the classroom”, *European Journal of Cognitive Psychology*, No. 19, pp. 494-513.

Marcoulides, G.A. and **R.E. Schumacker** (1996), *Advanced Structural Equation Modeling*, Erlbaum, Mahwah, New Jersey.

OECD (2013a), *PISA 2012 Results: Creative Problem Solving. Students’ Skills in Tackling Real-Life Problems* (Volume V), PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208070-en

OECD (2013b), *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed* (Volume II), PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201132-en

OECD (2009), *PISA 2006 Technical Report*, PISA, OECD Publishing, Paris, pp. 346-350.
http://dx.doi.org/10.1787/9789264048096-en

Paulhus, D. L., P. D. Harms, M. N. Bruce and **D. C. Lysy** (2003), “The over-claiming technique: Measuring self-enhancement independent of ability”, *Journal of Personality and Social Psychology*, Vol. 84(4), April 2003, pp. 890-904.
http://dx.doi.org/10.1037/0022-3514.84.4.890

Stark, S., O.S. Chernyshenko and **F. Drasgow** (2005), “An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model”, *Applied Psychological Measurement*, No. 29 (3), pp. 184-203.

Van de Gaer, E., P. Lietz and **R. Adams** (2011), “Application of the multivariate latent covariate model using PISA 2003”, Paper presented at the QEG Meeting, Melbourne (QEG(1109)2b), Australia, 26-28 September 2011.

Van Soest, A., L. Delaney, C. Harmon, A. Kapteyn and **J. P. Smith** (2011), “Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions”, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, pp. 174(3), pp. 575-595.

Weekley J.A. and **R.E. Ployhart** (2006), *Situational judgment tests: Theory, management, and application*, Erlbaum, Mahwah, New Jersey.

White, L.A. and **M.C. Young** (1998), “Development and validation of the Assessment of Individual Motivation (AIM)”, Paper presented at the Annual Meeting of the American Psychological Association, San Francisco.

Wothke, W. (2000), “Longitudinal and multi-group modeling with missing data”, in T.D. Little, K.U. Schnabel, and J. Baumert (eds.), *Modeling Longitudinal and Multiple Group Data: Practical Issues, Applied Approaches and Specific Examples*, Erlbaum, Mahwah, New Jersey. (Reproduced with permission).

Ziegler, M., C. MacCann and **R.D. Roberts** (2012), *New Perspectives on Faking in Personality Assessment*, Cambridge University Press, Oxford.

Zimmerman, J., P.K.Broder, J.J. Shaughnessy and **B.J. Underwood** (1977), “A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition”, *Intelligence*, No. 1, pp. 5-31.

18

Computer-Based Assessments

PISA 2012 included three computer-based assessments (CBA):

- problem solving (PS);
- mathematics (CBAM); and
- digital reading (DRA).

Problem solving was undertaken by 44 countries. The computer-based mathematical literacy assessment and the digital reading literacy assessment were combined into a single international option and this option was undertaken by 32 countries (see Table 18.1).

The number of locales in Table 18.1 refers to the number of unique country by language combinations that were actually administered. A locale refers to a country by language version of the test. For example, Australian English, Canadian English, Canadian French and French French are four locales in two languages spread across three countries. The number of locales presented in this table is slightly different from the one presented in Chapter 13 that only refers to locales that participated in the coding reliability exercise.

Table 18.1 **Number of countries and locales in PISA 2012 computer-based assessment**

	Number of countries	Number of locales
Problem solving only	12	16
Problem solving, mathematics and digital reading	32	51

Chapter 2 dealt with the associated test development activities, test design and framework coverage of the computer-based assessment. This chapter focuses on the technicalities and functionality of the delivery system and the various supporting systems.

ITEM RENDERING

Item authoring was a collaboration between three professional groups: assessment item writers, graphic artists and JavaScript® programmers. The activities of the assessment item writers are described in Chapter 2. Graphic artists illustrated the ideas of the writers, using a variety of images and where free drawing was required, a consistent colour palette and illustrative style. To aid readability and focus the viewer, the overall design layout was kept clean using minimal extraneous elements, dark colours and gradients that distract from relevant content.

The problem solving and computer-based mathematics items were authored for PISA 2012 in JavaScript®. The DRA items were first used in PISA 2009, when they were Flash based. For PISA 2012 the DRA items were regenerated in JavaScript®.

JavaScript™ enables the programmer to vary the appearance and position of HTML elements that are displayed in a browser. Images, text and other HTML elements can be moved, changed in colour or size, hidden or revealed in response to user input such as mouse dragging, clicking or data entry. With its ability to manipulate HTML elements, JavaScript® effectively becomes a relatively strong two-dimensional animation tool.

Many of the mathematics and problem-solving items developed for PISA 2012 required some form of animation. This included actions such as dragging around the bars on a graph, highlighting a route on a map or changing the text and colours of a simple-state machine.

The use of JavaScript® in PISA 2012 avoided the problems experienced with the Flash-based solution in 2009: the input of right to left languages was not supported; the input of Cyrillic script was not supported; and the input of characters with diacritics was not fully supported.

ONLINE ITEM REVIEW

The item review activities described in Chapter 2 were conducted using a secure online review system developed by the international contractor. Each National Centre was provided with one primary account to securely view, rate and comment upon each item. Several secondary accounts (as many as requested) were provided so that national experts could securely view, rate and comment upon each item. The primary account contained a reporting facility that enabled the National Project Manager to view the responses from national experts and collate these into a single response per country through the primary account.

TRANSLATION

Only English source versions of the items were released for translation (or adaptation in the case of English testing locales). The workflows of the translation and verification processes were facilitated with an online Translation Management System (TMS) developed by the international contractor.

Translation of material was achieved by first creating XML (Extensible Markup Language) Localisation Interchange File Format[1] (XLIFF) compliant text files, and then translating or editing these files in an XLIFF compatible translation editor. The translation editor was a modified version of the open source Open Language Tool (OLT) used in PISA 2009. DIPF with the input of cApStAn modified the OLT primarily to deal better with right to left languages, customise it to the PISA workflow, and eliminate major known bugs.

The Translation Management System supported the translation and linguistic quality assurance workflow described in Chapter 5.

Once translation was completed, national language versions of the test delivery software were provided to National Centres via downloads from the international contractor's FTP sites in both Australia and Germany.

SCHOOL COMPUTER REQUIREMENTS

The basic hardware requirement for delivering the test was the availability of a suitable computer for each student. To be suitable, a computer needed to satisfy the following three criteria:

- run on Windows®, XP® Windows Vista or Windows 7® operating system;
- have a colour display with resolution of at least 1024 x 768 pixels; and
- have at least one accessible USB port (e.g. at the front of the machine).

More specific requirements, including processing power, system memory and data transfer rates, were verified at the time of system diagnosis (for specifics, see the section on system diagnostic below).The delivery application and results storage were on a single USB flash drive. The delivery application had to be opened with administrative privileges.

The computers had to be located so that the test could be supervised by a single Test Administrator, and in such a manner that students could not easily observe each other's screens.

System diagnostic

To determine the suitability of a computer for delivering the computer-based assessment, a system diagnostic tool was distributed by the international contractor. The PISA 2012 Hardware Diagnostic was designed in part to emulate the test delivery system, but it also provided feedback on the computer's memory, processing power and screen resolution. The minimum requirements are detailed in Figure 18.1.

■ Figure 18.1 ■

Minimum requirements, as checked by the PISA 2012 system diagnostic

System area	Requirement for CBA
CPU speed installed	1 500 Mhz
Operating system	Windows XP®, Windows Vista® or Windows 7®
The memory installed on the computer in MB	Windows XP® = 512 MB Windows Vista® = 1 024 MB Windows 7® = 1 024 MB
Available memory	Optimal requirements: 358 MB for Windows XP® 717 MB for Windows Vista® 717 MB for Windows 7® -- Minimum requirements: 307 MB for Windows XP® 614 MB for Windows Vista® 614 MB for Windows 7®
Screen resolution	Minimal screen width = 1 024 px Minimal screen height = 768 px
Ensure Skype is not running on the computer	The CBA assessment cannot run with Skype™ running
The USB key data transfer rate	A minimum of 9 MB/s

As can be seen in Figure 18.1, the system diagnostic also checked that Skype™ was not running on the PC, as this interfered with the assessment display (Skype™ would recognise phone numbers in the assessment text and hyperlink them to Skype™ calling).

The system diagnostic also offered a virus check (of the USB, of the system memory, and of all available local and network drives). ClamWin™ Antivirus was used for this.

An optional advanced diagnostic check was also available. This check wrote a text file to the USB detailing as much relevant information as could be gathered about the system. The advanced check was useful where there was an issue preventing the optimal running of the assessment on a set of school computers but where all minimum requirements were met. The text file was sent to the international contractor for analysis to help determine the cause of any problem.

The system diagnostic interface source text could be localised – that is, replaced with text in the local language into which the assessment was translated.

TEST DELIVERY SYSTEM

The PISA 2012 computer-based assessments were delivered in schools via a set of software programs (described below) and national versions of the items, all bundled together onto a USB drive.

Generally, three variants of data collection were used by National Centres, sometimes in combination:

- the computers that existed in the sampled school were used to collect the data;
- laptops with preloaded software were carried into schools and used to collect the data; and
- students were transported to test centres.

An open-source computer-based assessment platform, TAO™ (*Testing Assisté par Ordinateur* or computer-based testing),[2] was used to sequence and store the items, store the results data, facilitate the student navigation, and provide all interface elements such as indicating progress through the test.

In PISA 2009 a Linux® OS based bootable system model was used to deliver a computer-based assessment. This model presented several problems, notably font recognition issues, drive recognition issues and the limitation of input methods to those recognised by Linux® (this was particularly frustrating for some Japanese and Chinese students). To overcome these problems in PISA 2012, the international contractor developed a Windows® OS based portable application.[3] The major advantages of the Windows application came from its use of the native operating system: hence there were no font recognition or rendering problems, no hardware driver recognition issues, and students could use their native input methods.

There were some disadvantages to using a Windows® application. Firstly, it could only be used on computers with a Windows® operating system, thereby excluding Apple® and Linux® based operating systems which in some countries were popular. Secondly, being an external application, the Windows® User Management System, which is embedded into all Windows® operating systems, required that the application was run with administrative privileges. This "administrative rights" issue imposed by Windows® was a significant constraint in some schools and systems, as it was not possible to circumvent the Windows® User Management System. Where the school's IT infrastructure is managed externally (e.g. by a contracted company or at the school district level), obtaining administrative permissions can be very difficult.

The Windows® application was basically a launcher for the test. In addition it was a launcher for the diagnostic tool that assessed the host computer for suitability for the test (see above section on system requirements).

Launching the test involved checking the computer for an available port, configuring a portable version of Firefox® to use that port, configuring the Apache™ web server and PHP (Hypertext Preprocessor) application server to run from the encrypted drive on the USB and then starting Apache™. Once Apache™ had started, the portable version of Firefox® was also started and opened to the entry page for the test. At this point the candidate was prompted to log in for the test to begin.

Access to the secure test was granted though a PHP login script. The student entered a 13-digit identifier (unique to each PISA student within a country). This identifier was then validated by entering a five-digit checksum. The checksum

was generated by applying a CRC (Cyclic Redundancy Check)16 security algorithm to the identifier. The identifier and checksum were communicated to the student via a form produced by the *KeyQuest* student sampling software. A second two-digit identifier (incorporating a checksum, also provided by *KeyQuest*) was entered by the student and this triggered automatic allocation of the appropriate pre-determined test form.

DATA CAPTURE AND SCORING STUDENT RESPONSES

Results data were written to the USB at regular intervals throughout the assessment. Data were transferred from National Centres to the international contractor via a secure FTP account.

In addition to results data that contributed directly towards students' cognitive scores, the system collected behavioural data such as time spent on task, sequence of pages visited, and use of stimulus elements such as drop down menus. In some cases, such behavioural data contributed to scoring.

All results data were stored in a single password protected ZIP file per student. Each ZIP file contained a set of XML files with each student's activities stored as a list of events. Of the 101 items across 3 domains (mathematics, problem solving and reading), 83 items were required to be automatically scored; each of these items was associated with a specific scoring algorithm. Event data were extracted from the ZIP files and processed with the following data being retained for further analysis: student details, item details, raw data log, response contents, score, time spent on each item. Auto-scored items include multiple choice, complex multiple choice, numeric input, dropdown select and many of the interactive tasks. The remaining open constructed-response items were imported into the Online Coding System, developed by the international contractor, to be coded by experts trained within each National Centre.

ONLINE CODING SYSTEM

The user interface for the Online Coding System was localisable – the language elements could be translated into any language variant used in the computer-based assessment. The National Project Manager was able to create user accounts with different roles:

- a coding supervisor responsible for organising and managing the online coding operation;
- leading coders, who played a central role in monitoring the quality of coding, as well as coding responses themselves; and
- coders.

The quality of the coding was monitored in several ways. For the more complex Constructed-Response items, multiple coding was applied to a subset of responses. Each multiply-coded response was coded by a minimum of four coders with each coder being allocated 100 responses for multiple coding when the assessment language was a major verified language, and 50 responses when the assessment language was a minor verified language. The remaining responses for these complex Constructed-Response items were assigned randomly to the coding group and coded once.

All responses to simple Constructed-Response items were coded once.

The coding assignment was handled by the online coding software, however the supervisor had the ability to determine a minimum proportion of responses to be allocated to each coder. This ensured that slow or absent coders still coded a reasonable number of responses, and that one or two fast coders did not code all the responses. Spreading the coding load among at least four coders reduced the risk of bias being introduced by an individual's harshness, leniency or misunderstanding of the coding guide. The method of reliability analysis and the results are described in detail in Chapter 13.

Another quality assurance measure was the introduction of control scripts for some items. These control scripts were translations of standard responses, to which the code to be applied was known to the international contractor. Control scripts were inserted into the coding workflow, blind to the coder, and analysed by the international contractor to determine whether codes had been applied correctly and consistently across countries. The method and results of this analysis are described in Chapter 13.

During the coding of the item, the Leading Coder spot-checked the work of coders each day. Spot-checking involved a review of codes assigned to responses. A general guide was that at least 10 responses per item were spot-checked.

If a coder was uncertain about the code to assign to a particular response, the response could be marked for review and it would be sent automatically to a leading coder or supervisor for advice.

The Online Coding System provided several reports to help the coding supervisor manage the quality and workflow of the coding process, including:

- a short summary of allocated items listing the coders assigned to that item, the number of responses remaining for each coder, the number of single-coded responses that have yet to be assigned and the total number of responses coded by each coder;
- a complete summary of allocated items which, in addition to the information in the short summary, contained data for each item on the minimum and maximum number of single-coded responses assigned to each coder, the total number of responses that had been multiple-marked and single-marked, and, when coding was complete for each item, item reliability statistics were made available; and
- a report listing any instances in which the missing code had been applied inconsistently to the same response.

Notes

1. For a description, see *http://en.wikipedia.org/wiki/XLIFF*. Version 1.1 was current and specifications for this version can be found at *http://www.oasis-open.org/committees/xliff/documents/cs-xliff-core-1.1-20031031.htm*.

2. Developed by the Centre de Recherche Public (CRP): Henri Tudor and the Université du Luxembourg. See *https://www.tao.lu/*.

3. Developed by the Deutsche Institut für Internationale Pädagogische Forschung (DIPF) in co-operation with the Centre de Recherche Public (CRP).

19

International Databases

This chapter describes the three databases containing the PISA 2012 international data. The PISA 2012 International Database is described first. This description is followed by shorter descriptions of the computer-based assessment database and the financial literacy database, highlighting the differences between them and the PISA 2012 international database.

FILES IN THE DATABASE

The PISA 2012 international database consists of five data files: three with student responses, one with school responses and one with parent responses. All are provided in fixed width text (or ASCII) format with the corresponding SAS® and SPSS® control files.

Student files

The student performance and questionnaire data file (*filename: INT_STU12_DEC03.txt*, available at *www.oecd.org/pisa*) contains, for each student who participated in the assessment, the following information:

- Identification variables for the country, school and student.
- The student responses to the three questionnaires, *i.e.*, the student questionnaire, the information communication technology international option questionnaire, and the education career international option questionnaire.
- The indices derived from each student's responses to the original questions in the questionnaires.
- The students' performance scores in mathematics, reading, science, and the seven subscales of mathematics (five plausible values for each of these domains).
- The student weight variable and 80 Fay's replicates for the computation of the sampling variance estimates and a senate weight.
- A normalised (senate) weight variable for analyses of student performance across a group of countries where contributions from each of the countries in the analysis are desired to be equal regardless of their population size. The senate weight makes the population of each country to be 1000 to ensure an equal contribution by each of the countries in the analysis. This weight is only applicable to the student performance scores (plausible values) that do not contain missing values. Its application to other variables might be compromised by its dependence on the patterns of missing data.
- Three sampling related variables: the randomised final variance stratum, the final variance unit and the original explicit strata, mostly labelled by country.
- A test language variable from the cognitive test.
- A database version identifier with the date of the release.

Three sets of indices are provided in the student questionnaire files. The first set is based on a transformation of one variable or it is based on a combination of information gathered from two or more variables. Forty-nine indices of the first type are included in the database. The second set is the result of a Rasch scaling and consists of Weighted Likelihood Estimate indices. Twenty-four indices from the Student Questionnaire and seven indices from the information communication technology questionnaire are included in the database from this second type. The third set is the result of applying an anchoring vignettes approach in the preparation of indices and a subsequent Rasch scaling of the indices, and consists of weighted likelihood estimate indices (see Chapter 16). Thirteen indices of this type are included in the database. The *PISA index of economic, social and cultural status* (ESCS) is derived as factor scores from a principal component analysis and is also included in the database. For a full description of the indices see Chapter 16.

For each domain, i.e. mathematics, reading and science, and for each subscale in mathematics, i.e. the four content categories of *change and relationships, quantity, space and shape, uncertainty and data,* as well as the three process categories of *employ, formulate* and *interpret,* a set of five plausible values (transformed to the PISA scale) are provided.

It is important to note that four content scales and three process scales are based on the same test items. As such, it is inappropriate to jointly analyse any of the four content scales with any of the three process scales. For example, it would not be meaningful to correlate or otherwise compare performance on the *space and shape* scale, with performance on the *employ* scale as some of the items are included in both of these two scales.

The metrics of the performance scales are established so, that in the year that the scale is first established, the OECD students' mean score is 500 and the pooled OECD standard deviation is 100. The reading scale was established in

2000, the mathematics scale in 2003 and the science scale in 2006. When establishing the scale the data are weighted to ensure that each OECD adjudicated country is given equal weight.

Plausible values for reading were mapped to the PISA 2000 scale, plausible values for mathematics were mapped to the PISA 2003 scale and plausible values for science were mapped to the PISA 2006 scale. See Chapter 12 for details of these mappings.

The variable *W_FSTUWT* is the final student weight. The sum of these weights constitutes an estimate of the size of the target population. When analysing weighted data at the international level, large countries have a greater contribution to the results than small countries. This weighting is used for the OECD total in the tables of the international report for the first results from PISA 2012 (OECD, 2014). To weight all countries equally for a summary statistic, the OECD average is computed and reported. The OECD average is computed as follows. First, the statistic of interest is computed for each OECD country using the final student weights. Second, the mean of the country statistics is computed and reported as the OECD average.[1]

For a full description of the weighting methodology and the calculation of the weights, see Chapter 8. How to use weights in analyses of the database is described in detail in the *PISA Data Analysis Manual* for SPSS® or SAS® users (OECD, 2009a, 2009b),[2] which is available at *www.oecd.org/pisa/pisaproducts/*. The data analysis manual also explains the theory behind sampling, plausible values and replication methodology and how to compute standard errors in the case of two-stage, stratified sampling designs.

Two versions of the student cognitive files are available:

1) a version that contains single-digit and original responses (*filename: INT_COG12_DEC03.txt*, available at *www.oecd.org/pisa*) and
2) a version that contains scored responses (*filename: INT_COG12_S_DEC03.txt*, available at *www.oecd.org/pisa*)

For each student who participated in the assessment, the following information is available:

- Identification variables for the country, school and student.
- Test booklet identification.
- The student responses to the cognitive items. When the original responses consist of multiple digits (complex multiple choice or open ended items), the multiple digits were recoded into single-digit variables for use in scaling software). A "T" was added to the end of the recoded single digit variable names. The original response variables have been added at the end of the single-digit, unscored file (with an "R" at the end of the variable name see further below). For the double-digit variables (*PM155Q02, PM155Q03, PM462Q01, PS131Q02, PS131Q04, PS269Q03, PS438Q03*) a "D" was added to the end of the recoded single-digit variable.
- Test language.
- Database version with the date of the release.

The PISA items are organised into units. Each unit consists of a stimulus (consisting of a piece of text or related texts, pictures or graphs) followed by one or more questions. A unit is identified by a short label and by a long label. The units' short labels consist of five characters and form the first part of the variable names in the data files. The first two characters are PR, PM or PS for reading, mathematics or science respectively in the pencil and paper component of the assessment. The next three characters indicate the unit within the domain. For example, PM155 is a paper and pencil mathematics unit. The item names (usually eight or nine digits) represent questions within a unit and are used as variable names (in the current example the variable names for items within the unit are *PM155Q01, PM155Q02D, PM155Q03D* and *PM155Q04T*). Thus items within a unit have the same initial five characters plus a question number. Responses that needed to be recoded into single-digit variables have a "T" or "D" at the end of the variable name. The original multiple-digit responses have been added to the end of the single-digit and original responses file (*filename: INT_COG_DEC03.txt*) with an "R" at the end of the variable name (for example, the variable *PM155Q02D* is a recoded item with the corresponding original responses in *PM155Q02R* at the end of the file). The full variable label indicates the domain the unit belongs to, the PISA cycle in which the item was first used, the full name of the unit and the question number. For example, the variable label for PM155Q01 is *'MATH - P2000 POPULATION PYRAMIDS (Q01)'*. The variable name for this item was *M155Q01* in PISA 2000-PISA 2009. Following the extended naming convention of PISA 2012, the variable label was modified to *PM155Q01* to reflect that it belongs to a set of items for the paper-based assessment.

The scored data file (*filename: INT_COG12_S_DEC03.txt*) only includes one single-digit variable per item with scores instead of response categories.

In both files, the cognitive items are sorted by domain and alphabetically by item name within domain. This means that the mathematics items appear at the beginning of the file, followed by the reading items and then the science items. Within domains, units with smaller numeric identification appear before those with larger identification, and within each unit, the first question will precede the second, and so on.

School file

The school questionnaire data file (*filename: INT_SCQ12_DEC03.txt*, available at *www.oecd.org/pisa*) contains for each school that participated in the assessment, the following information:

- The identification variables for the country and school.
- The school responses on the School Questionnaire.
- The school indices derived from the original questions in the School Questionnaire.
- The school weight and senate school weight.
- Explicit strata with national labels.
- Database version with the date of the release.

The school file contains the original variables collected through the school context questionnaire. In addition, two types of indices are provided in the School Questionnaire files. The first set is based on a transformation of one variable or on a combination of two or more variables. The database includes twenty one indices from this first type. The second set is the result of a Rasch scaling and consists of weighted likelihood estimate indices. Twelve indices are included in the database from this second type. For a full description of the indices and how to interpret them, see Chapter 16. The school weight (*W_FSCHWT*) is the trimmed school base weight adjusted for non-response (see also Chapter 8).

Although the student samples were drawn from within a sample of schools, the school sample was designed to optimise the resulting sample of students, rather than to give an optimal sample of schools. For this reason, it is always preferable to analyse the school-level variables as attributes of students, rather than as elements in their own right (Gonzalez and Kennedy, 2003).

Following this recommendation one would not estimate the percentages of private schools versus public schools, for example, but rather the percentages of students attending a private school or public schools. From a practical point of view, this means that the school data should be merged with the student data file prior to analysis.

For general information about analyses of the data, see the *PISA Data Analysis Manual* for SPSS® or SAS® users (OECD, 2009a, 2009b),[3] also available at *www.oecd.org/pisa/pisaproducts/*. Chapter 10 of the *PISA Data Analysis Manual* describes analyses with school level variables. Chapter 15 is about multi-level analysis using PISA data.

Parent file

The parent questionnaire file (*filename: INT_PAQ12_DEC03.txt*, available at *www.oecd.org/pisa*) contains the following information:

- identification variables for the country, school and student;
- the parents' responses to the parent questionnaire;
- the parent indices derived from the original questions in the parent questionnaire; and
- Database version with the date of the release.

The parent file contains the original variables collected through the Parent Context Questionnaire as a national option instrument. In addition, two types of indices are provided in the Parent Questionnaire file. The first set is based on a transformation of one variable or on a combination of two or more variables. The database includes twenty three indices from this first type. The second set is the result of a Rasch scaling and consists of Weighted Likelihood Estimate indices. Five indices are included in the database from this second type. For a detailed description of the indices see Chapter 16.

Due to the high parent non-response in most countries, caution is needed when analysing these data. Non-response is unlikely to be random. When using the final student weights from the student file, the weights of valid students in the analysis do not sum to the population size of parents of PISA eligible students. A weight adjustment is not provided in the database.

RECORDS IN THE DATABASE

Records included in the database

In the student and parent files

- All PISA students who attended paper-based or computer-based test sessions.
- PISA students who only attended the questionnaire session are included if they provided at least one response to the Student Questionnaire and the father's or the mother's occupation is known from the Student or the Parent Questionnaire.

In the school file

- All participating schools – that is, any school where at least 25% of the sampled eligible, non-excluded students were assessed – have a record in the school-level international database, regardless of whether the school returned the School Questionnaire.

Records excluded from the database

Student and parent file

- Additional data collected by countries as part of national or international options.
- Sampled students who were reported as not eligible, students who were no longer at school, students who were excluded for physical, mental or linguistic reasons, and students who were absent on the testing day.
- Students who refused to participate in the assessment sessions.
- Students from schools where less than 25% of the sampled and eligible, non-excluded students participated.

School file

- Additional data collected by countries as part of national or international options.
- Schools where fewer than 25% of the sampled eligible, non-excluded students participated in the testing sessions.

REPRESENTING MISSING DATA

The coding of the data distinguishes between four different types of missing data:

- Item level non-response: 9 for a one-digit variable, 99 for a two-digit variable, 999 for a three-digit variable, and so on. Missing codes are shown in the codebooks. This missing code is used if the student or school principal was expected to answer a question, but no response was actually provided.
- Multiple or invalid responses: 8 for a one-digit variable, 98 for a two-digit variable, 998 for a three-digit variable, and so on. For the multiple-choice items code 8 is used when the student selected more than one of the answer options.
- Not-administered: 7 for a one-digit variable, 97 for a two-digit variables, 997 for a three-digit variable, and so on. Generally this code is used for cognitive and questionnaire items that were not administered to the students as a result of the balanced incomplete block test design used in PISA, and for items that were deleted after assessment because of misprints or translation errors.
- Not reached items: all consecutive missing values clustered at the end of test session were replaced by the non-reached code, "r", except for the first value of the missing series, which is coded as item level non-response (code 9).

HOW ARE STUDENTS AND SCHOOLS IDENTIFIED?

The student identification from the student and parent files consists of three variables, which together form a unique identifier for each student:

- A country identification variable *CNT*. The values for this variable are drawn from the ISO 3166-1 ALPHA-3 classification (*http://unstats.un.org/unsd/methods/m49/m49alpha.htm*) used by the United Nations. Note that for several PISA 2012 participants the value for the *CNT* variable does not correspond to this classification system. This occurs for two

possible reasons. Firstly, where a National Centre represents only part of the country. The codes of this type are QCN for Shanghai representing part of China, QCY for part of Cyprus, QRS for the Perm region of the Russian Federation, QUA, QUB and QUC for the three states of the United States, Florida, Connecticut and Massachusetts correspondingly. Secondly, where the National Centre represented only part of the country in a previous cycle, and even though the full country is participating in PISA 2012, the *CNT* value has been preserved for consistency. The only participant's code of this type is ARE for the United Arab Emirates.

- A school identification variable labelled *SCHOOLID*.
- A student identification variable labelled *STIDSTD*.

The school identification consists of two variables, which together form a unique identifier for each school:

- The country identification variable labelled *CNT*. The country codes used in PISA are the ISO numerical three-letter country codes.
- The school identification variable labelled *SCHOOLID*.

Some additional identification variables are included in all data files of the database as follows:

A variable *SUBNATIO* has been included to differentiate adjudicated sub-national entities within countries. This variable (*SUBNATIO*) is used as follows:

- *Belgium*. The value "0560100" is assigned to the Flemish region and "0560000" to the French and German regions of Belgium
- *Spain*. The value "7240100" is assigned to Andalusia, "7240200" to Aragon, "7240300" to Asturias, "7240400" to "Balearic Islands", "7240600" to Cantabria, "7240700" to Castile and Leon, "7240900" to Catalonia, "7241000" to Extremadura, "7241100" to Galicia, "7241200" to La Rioja, "7241300" to Madrid, "7241400" to Murcia, "7241500" to Navarre, and "7241600" to Basque Country. The value "7240000" is assigned to the rest of the country.
- *United Kingdom*. The value "8260000" is assigned to England, Northern Ireland and Wales and the value "8262000" is assigned to Scotland.
- *Argentina*. The value of "0320100" is assigned to the Autonomous City of Buenos Aires. The value "0320000" is assigned to the rest of the country.
- *United Arab Emirates*. The value of "7840100" is assigned to Abu Dhabi and the value of "7840200" is assigned to Dubai. The value "7840000" is assigned to the rest of the country.
- *Perm region of the Russian Federation*. The value of "6430059" is assigned to Perm data which was collected separately from the Russian Federation data.
- *Florida, Connecticut and Massachusetts of the United States*. The value "8400100" is assigned to Florida, the value "8400200" is assigned to Connecticut and the value "8400300" is assigned to Massachusetts.

A variable *NC* is included to identify National Centres. The variable is based on the ISO numerical three-digit country codes (*http://unstats.un.org/unsd/methods/m49/m49alpha.htm*) with an addition of leading zero and a trailing double-digit code having value of "*00*" for the majority of the countries and a different value for those countries where administration of PISA 2012 was performed by separate National Centres or where not the whole country participated in the assessment. For example, the *NC* code for Australia is "*003600*". *NC* code of Shanghai representing a part of China is "*015601*" and *NC* code for Scotland where PISA 2012 was administered separately from the rest of the United Kingdom is "*082620*".

A variable *STRATUM* is also included to differentiate sampling strata. The variable is created as a concatenation of a 3-letter country code and a two-digit region identifier and two-digit original stratum identifier. Value labels are provided in the control files to indicate the population defined by each stratum.[4]

COMPUTER-BASED ASSESSMENT DATABASE

For the 44 countries that participated in PISA 2012 problem solving or problem solving and computer-based assessment of mathematics and reading literacy a separate database was prepared.

With the exception of Brazil, Italy and Spain the number of cases included in the computer-based assessment (CBA) database is the same as the number of cases in the PISA 2012 international database. Brazil, Italy and Spain chose to

subsample schools from their large national school sample — see Chapter 4 for details of CBA sampling. The weight and replicate weight variables for these three countries have been adjusted in the CBA database to reflect this subsampling. For all other countries, the CBA weights and paper-based weights are identical.

The PISA CBA database consists of five data files: three with student responses, one with school responses and one with parent responses. All are provided in fixed-width text (or ASCII) format with the corresponding *SAS®* and *SPSS®* control files.

Student files

Student performance and questionnaire data file (*filename: CBA_STU12_MAR31.txt*, available at *www.oecd.org/pisa*).

For each student all the variables that are included in the international database are also included in CBA data file. The following additional information is also included:

- The students' performance scores in problem solving (five plausible values), computer-based mathematics and digital reading (five plausible values for each domain). It is necessary to note that 32 countries participated in both problem solving and computer-based assessment of literacy and 12 countries in problem-solving assessment only. Thus for 12 countries, performance scores in computer-based mathematics and digital reading contain missing values for all students.
- CBA Language variable.
- CBA Test Form.

Two versions of the student cognitive files are available:

1) A version that contains single-digit and original responses (*filename: CBA_COG12_MAR31.txt*, available at *www.oecd.org/pisa*) and
2) a version that contains scored responses (*filename: CBA_COG12_S_MAR31.txt*, available at *www.oecd.org/pisa*)

Additional information included in the CBA cognitive files is as follows:

- Original and coded responses for CBA items
- CBA Language variable
- CBA Test Form

Parent file

The Parent Questionnaire data file (*filename: CBA_PAQ12_MAR31.txt*, available at *www.oecd.org/pisa*).

The CBA parent file contains the same information as the international data file for the participating countries.

School file

The School Questionnaire data file (*filename: CBA_SCQ12_MAR31.txt*, available at *www.oecd.org/pisa*).

The CBA school file contains the same information as the international data file for the participating countries.

FINANCIAL LITERACY DATABASE

For the 18 countries that participated in the optional PISA 2012 financial literacy assessment a separate database was prepared.

The sampling of students for the financial literacy assessment was performed in such a way that there was no overlap with the main international database. Thus students reported in this database have no common records with the students included in international database – see Chapter 4 for details of the financial literacy sampling. The weight and replicate weight variables were calculated for each country along with the weights for students in the international database with the only exception of Spain where a special file was created for financial literacy students.

The PISA financial literacy database consists of five data files: three with student responses, one with school responses and one with parent responses. All are provided in fixed-width text (or ASCII) format with the corresponding SAS® and SPSS® control files.

Student files

Student performance and questionnaire data file (*filename: FIN_STU12_MAR31.txt*, available at *www.oecd.org/pisa*)

For each student who participated in financial literacy assessment all the variables that are included in the international database are also included in financial literacy data file. The following variables are included in addition to (or instead of) the variables in the international database:

- Instead of performance scores in mathematics, reading and science reported in the international database three sets of plausible values calculated using only data from students who participated in the financial literacy assessment. Those sets are performance scores in mathematics, reading and financial literacy (five plausible values for each domain).
- Sixteen questionnaire items added to the Student Questionnaire for the countries that participated in the financial literacy assessment.

Two versions of the student cognitive files are available:

- a version that contains single-digit and original responses (*filename: FIN_COG12_MAR31.txt*, available at *www.oecd.org/pisa*) and
- a version that contains scored responses (*filename: FIN_COG12_S_MAR31.txt*, available at *www.oecd.org/pisa*)

Additional information included in the financial literacy cognitive files is as following:

- Original and coded responses for financial literacy items.

Parent file

The Parent Questionnaire data file (*filename: FIN_PAQ12_MAR31.txt*, available at *www.oecd.org/pisa*)

The financial literacy parent file contains the same information as the international data file for the participating countries.

School file

The School Questionnaire data file (*filename: FIN_SCQ12_MAR31.txt*, available at *www.oecd.org/pisa*)

The financial literacy school file contains the same information as the international data file for the participating countries.

Further information

A full description on how to analyse the PISA database in accordance with the complex methodologies used to collect and process the data is provided in the *PISA Data Analysis Manual* (OECD, 2009),[5] available at *www.pisa.oecd.org*.

Notes

1. The definition of the OECD average has changed between PISA 2003 and PISA 2006. In PISA 2000 and 2003, the OECD average was based on a pooled, equally weighted database. To compute the OECD average the data was weighted by an adjusted student weight variable that made the sum of the weights equal in all countries.

2. This publication is focused on PISA 2006, but the principles remain the same for PISA 2012.

3. This publication is focused on PISA 2006, but the principles remain the same for PISA 2012.

4. Note that not all participants permit the identification of all sampling strata in the database.

5. This publication is focused on PISA 2006, but the principles remain the same for PISA 2012.

References

Gonzalez, E.J. and **A.M. Kennedy** (2003), *PIRLS 2001 User Guide for the International Database*, Boston College, Chestnut Hill.

OECD (2014), *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*, PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208780-en

OECD (2009a), *PISA Data Analysis Manual: SPSS, Second Edition*, PISA, OECD Publishing, Paris.

OECD (2009b), *PISA Data Analysis Manual: SAS, Second Edition*, PISA, OECD Publishing, Paris.

Annexes

ANNEX A – MAIN SURVEY ITEM POOL CLASSIFICATION

[Part 1/4]

Table A.1 PISA 2012 Main Survey mathematics item classification

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Content
PM00FQ01	"MATH - P2012 Apartment Purchase Q1"	Italy	Italian	Constructed Response Expert	Space and shape
PM00GQ01	"MATH - P2012 An Advertising Column Q1"	Czech	English	Constructed Response Manual	Space and shape
PM00KQ02	"MATH - P2012 Wheelchair Basketball Q2"	Canada	English	Constructed Response Expert	Space and shape
PM033Q01	"MATH - P2000 A View with a Room Q1"	CITO	Dutch	Simple Multiple Choice	Space and Shape
PM034Q01T	"MATH - P2000 Bricks Q1"	CITO	Dutch	Constructed Response Auto-coded	Space and Shape
PM155Q01	"MATH - P2000 Pop Pyramids Q1"	CITO	Dutch	Constructed Response Expert	Change and Relationships
PM155Q02D	"MATH - P2000 Pop Pyramids Q2"	CITO	Dutch	Constructed Response Expert	Change and Relationships
PM155Q03D	"MATH - P2000 Pop Pyramids Q3"	CITO	Dutch	Constructed Response Expert	Change and Relationships
PM155Q04T	"MATH - P2000 Pop Pyramids Q4"	CITO	Dutch	Complex Multiple Choice	Change and Relationships
PM192Q01T	"MATH - P2000 Containers Q1"	Germany	German	Complex Multiple Choice	Change and Relationships
PM273Q01T	"MATH - P2000 Pipelines Q1"	Czech Republic	Czech	Complex Multiple Choice	Space and Shape
PM305Q01	"MATH - P2000 Map Q1"	ACER	English	Simple Multiple Choice	Space and Shape
PM406Q01	"MATH - P2003 Running Tracks Q1"	ACER	English	Constructed Response Expert	Space and Shape
PM406Q02	"MATH - P2003 Running Tracks Q2"	ACER	English	Constructed Response Expert	Space and Shape
PM408Q01T	"MATH - P2003 Lotteries Q1"	ACER	English	Complex Multiple Choice	Uncertainty and data
PM411Q01	"MATH - P2003 Diving Q1"	ACER	English	Constructed Response Manual	Quantity
PM411Q02	"MATH - P2003 Diving Q2"	ACER	English	Simple Multiple Choice	Uncertainty and data
PM420Q01T	"MATH - P2003 Transport Q1"	ACER	English	Complex Multiple Choice	Uncertainty and data
PM423Q01	"MATH - P2003 Tossing Coins Q1"	ACER	English	Simple Multiple Choice	Uncertainty and data
PM442Q02	"MATH - P2003 Braille Q2"	ACER	English	Constructed Response Manual	Quantity
PM446Q01	"MATH - P2003 The Thermometer Cricket Q1"	ACER	English	Constructed Response Manual	Change and Relationships
PM446Q02	"MATH - P2003 The Thermometer Cricket Q2"	ACER	English	Constructed Response Expert	Change and Relationships
PM447Q01	"MATH - P2003 Tile Arrangement Q1"	ACER	English	Simple Multiple Choice	Space and Shape
PM462Q01D	"MATH - P2003 The Third Side Q1"	Sweden	English	Constructed Response Expert	Space and Shape
PM464Q01T	"MATH - P2003 The Fence Q1"	Sweden	English	Constructed Response Auto-coded	Space and Shape
PM474Q01	"MATH - P2003 Running Time Q1"	Canada	English	Constructed Response Manual	Quantity
PM496Q01T	"MATH - P2003 Cash Withdrawal Q1"	ACER	English	Complex Multiple Choice	Quantity
PM496Q02	"MATH - P2003 Cash Withdrawal Q2"	ACER	English	Constructed Response Manual	Quantity
PM559Q01	"MATH - P2003 Telephone Rates Q1"	Italy	English	Simple Multiple Choice	Quantity
PM564Q01	"MATH - P2003 Chair Lift Q1"	Italy	English	Simple Multiple Choice	Quantity
PM564Q02	"MATH - P2003 Chair Lift Q2"	Italy	English	Simple Multiple Choice	Uncertainty and data
PM571Q01	"MATH - P2003 Stop the Car Q1"	Germany	German	Simple Multiple Choice	Change and Relationships
PM603Q01T	"MATH - P2003 Number Check Q1"	Austria	German	Complex Multiple Choice	Quantity
PM800Q01	"MATH - P2003 Computer Game Q1"	Canada	English	Simple Multiple Choice	Quantity
PM803Q01T	"MATH - P2003 Labels Q1"	Canada	English	Constructed Response Auto-coded	Uncertainty and data
PM828Q01	"MATH - P2003 Carbon Dioxide Q1"	Netherlands	English	Constructed Response Expert	Change and Relationships
PM828Q02	"MATH - P2003 Carbon Dioxide Q2"	Netherlands	English	Constructed Response Manual	Uncertainty and data
PM828Q03	"MATH - P2003 Carbon Dioxide Q3"	Netherlands	English	Constructed Response Manual	Quantity
PM903Q01	"MATH - P2012 Drip Rate Q1"	ACER	English	Constructed Response Expert	Change and relationships
PM903Q03	"MATH - P2012 Drip Rate Q3"	ACER	English	Constructed Response Manual	Change and relationships
PM905Q01T	"MATH - P2012 Tennis Balls Q1"	ACER	English	Complex Multiple Choice	Quantity
PM905Q02	"MATH - P2012 Tennis Balls Q2"	ACER	English	Constructed Response Expert	Quantity
PM906Q01	"MATH - P2012 Crazy Ants Q1"	ACER	English	Simple Multiple Choice	Quantity
PM906Q02	"MATH - P2012 Crazy Ants Q2"	ACER	English	Constructed Response Expert	Quantity
PM909Q01	"MATH - P2012 Speeding Fines Q1"	aSPe	English	Constructed Response Manual	Quantity
PM909Q02	"MATH - P2012 Speeding Fines Q2"	aSPe	English	Simple Multiple Choice	Quantity
PM909Q03	"MATH - P2012 Speeding Fines Q3"	aSPe	English	Constructed Response Expert	Change and relationships
PM915Q01	"MATH - P2012 Carbon Tax Q1"	ILS	English	Simple Multiple Choice	Uncertainty and data
PM915Q02	"MATH - P2012 Carbon Tax Q2"	ILS	English	Constructed Response Manual	Change and relationships
PM918Q01	"MATH - P2012 Charts Q1"	IPN/Kassel	English	Simple Multiple Choice	Uncertainty and data
PM918Q02	"MATH - P2012 Charts Q2"	IPN/Kassel	English	Simple Multiple Choice	Uncertainty and data
PM918Q05	"MATH - P2012 Charts Q5"	IPN/Kassel	English	Simple Multiple Choice	Uncertainty and data
PM919Q01	"MATH - P2012 Zs Fan Merchandise Q1"	IPN/Kassel	English	Constructed Response Manual	Quantity
PM919Q02	"MATH - P2012 Zs Fan Merchandise Q2"	IPN/Kassel	English	Constructed Response Manual	Quantity
PM923Q01	"MATH - P2012 Sailing Ships Q1"	IPN/Kassel	English	Simple Multiple Choice	Quantity
PM923Q03	"MATH - P2012 Sailing Ships Q3"	IPN/Kassel	English	Simple Multiple Choice	Space and shape
PM923Q04	"MATH - P2012 Sailing Ships Q4"	IPN/Kassel	English	Constructed Response Expert	Change and relationships
PM924Q02	"MATH - P2012 Sauce Q2"	NIER	English	Constructed Response Manual	Quantity
PM934Q01	"MATH - P2012 London Eye Q1"	Mathematics Expert Group	English	Constructed Response Manual	Space and shape

[Part 2/4]

Table A.1 PISA 2012 Main Survey mathematics item classification

Unit Item Code	Context	Process	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
						Delta	Tau(1)	Tau(2)	1	2
PM00FQ01	Personal	Formulate	PM6A	44.64	(0.24)	0.39027			576.2	
PM00GQ01	Personal	Formulate	PM5	8.78	(0.13)	2.75209			760.1	
PM00KQ02	Personal	Formulate	PM4	14.85	(0.18)	1.97967			699.9	
PM033Q01	Personal	Interpret	PM1	75.78	(0.20)	-1.44130			433.5	
PM034Q01T	Occupational	Formulate	PM1	42.38	(0.23)	0.42603			579.0	
PM155Q01	Scientific	Interpret	PM1	67.67	(0.22)	-0.84340			480.0	
PM155Q02D	Scientific	Employ	PM1	61.57	(0.21)	-0.44941	0.74491	-0.74491	492.4	529.1
PM155Q03D	Scientific	Employ	PM1	18.68	(0.17)	1.56865	-1.56865	1.56865	639.4	696.6
PM155Q04T	Scientific	Interpret	PM1	55.87	(0.23)	-0.38438			515.8	
PM192Q01T	Scientific	Formulate	PM2	42.44	(0.24)	0.44066			580.0	
PM273Q01T	Occupational	Employ	PM3	51.46	(0.23)	-0.30936			521.6	
PM305Q01	Societal	Employ	PM2	60.36	(0.22)	-0.70057			491.2	
PM406Q01	Societal	Employ	PM2	25.62	(0.22)	1.26690			644.5	
PM406Q02	Societal	Formulate	PM2	16.89	(0.19)	1.92572			695.8	
PM408Q01T	Societal	Interpret	PM3	39.39	(0.23)	0.57862			590.9	
PM411Q01	Societal	Employ	PM1	51.08	(0.25)	0.00939			546.6	
PM411Q02	Societal	Interpret	PM1	45.71	(0.24)	0.03440			548.4	
PM420Q01T	Personal	Interpret	PM3	50.02	(0.23)	-0.08922			538.8	
PM423Q01	Personal	Interpret	PM2	79.05	(0.19)	-1.85314			401.4	
PM442Q02	Societal	Interpret	PM1	38.26	(0.24)	0.60225			592.7	
PM446Q01	Scientific	Formulate	PM3	68.57	(0.23)	-0.95034			471.8	
PM446Q02	Scientific	Formulate	PM3	6.82	(0.13)	2.97548			777.6	
PM447Q01	Societal	Employ	PM3	68.33	(0.22)	-1.03098			465.5	
PM462Q01D	Scientific	Employ	PM1	12.20	(0.16)	1.70520	0.65379	-0.65379	658.6	698.6
PM464Q01T	Societal	Formulate	PM3	23.66	(0.21)	1.35931			651.6	
PM474Q01	Personal	Employ	PM1	74.31	(0.20)	-1.32556			442.6	
PM496Q01T	Societal	Formulate	PM2	52.97	(0.24)	-0.30933			521.6	
PM496Q02	Societal	Employ	PM2	66.74	(0.22)	-0.97302			470.0	
PM559Q01	Societal	Interpret	PM3	63.14	(0.23)	-0.95819			471.2	
PM564Q01	Societal	Formulate	PM2	46.11	(0.24)	-0.07690			539.8	
PM564Q02	Societal	Formulate	PM2	45.82	(0.24)	-0.03745			542.9	
PM571Q01	Scientific	Interpret	PM2	47.67	(0.24)	0.04707			549.4	
PM603Q01T	Scientific	Employ	PM2	45.07	(0.23)	0.19753			561.1	
PM800Q01	Personal	Employ	PM3,PMUH	88.39	(0.14)	-2.93302			317.3	
PM803Q01T	Occupational	Formulate	PM1	29.18	(0.22)	1.25570			643.6	
PM828Q01	Scientific	Employ	PM3	28.45	(0.21)	0.98562			622.5	
PM828Q02	Scientific	Employ	PM3	55.95	(0.23)	-0.44691			511.0	
PM828Q03	Scientific	Employ	PM3	28.02	(0.21)	0.96819			621.2	
PM903Q01	Occupational	Employ	PM6A	22.23	(0.19)	1.13372	0.48693	-0.48693	610.5	657.7
PM903Q03	Occupational	Employ	PM6A	25.70	(0.21)	1.10252			631.7	
PM905Q01T	Occupational	Interpret	PM7A	77.71	(0.19)	-1.62564			419.1	
PM905Q02	Occupational	Interpret	PM7A	50.05	(0.25)	-0.02564			543.8	
PM906Q01	Scientific	Employ	PM4	60.65	(0.24)	-0.65141			495.1	
PM906Q02	Scientific	Employ	PM4	42.12	(0.24)	0.12598	0.89486	-0.89486	539.8	571.4
PM909Q01	Societal	Interpret	PM5,PMUH	89.34	(0.14)	-2.56225			346.1	
PM909Q02	Societal	Employ	PM5,PMUH	63.12	(0.23)	-0.69123			491.9	
PM909Q03	Societal	Interpret	PM5,PMUH	35.70	(0.23)	0.76662			605.5	
PM915Q01	Societal	Employ	PM4,PMUH	40.18	(0.23)	0.38327			575.6	
PM915Q02	Societal	Employ	PM4,PMUH	68.24	(0.23)	-1.20630			451.8	
PM918Q01	Societal	Interpret	PM6A	87.27	(0.15)	-2.54294			347.7	
PM918Q02	Societal	Interpret	PM6A	79.54	(0.19)	-1.67876			415.0	
PM918Q05	Societal	Employ	PM6A	76.67	(0.20)	-1.50962			428.2	
PM919Q01	Personal	Employ	PM7A	84.51	(0.17)	-2.04272			386.6	
PM919Q02	Personal	Formulate	PM7A	44.72	(0.23)	0.21442			562.4	
PM923Q01	Scientific	Employ	PM6A	59.49	(0.23)	-0.43663			511.7	
PM923Q03	Scientific	Employ	PM6A	49.79	(0.23)	-0.09349			538.5	
PM923Q04	Scientific	Formulate	PM6A	15.28	(0.17)	2.00671			702.1	
PM924Q02	Personal	Formulate	PM6A	63.45	(0.23)	-0.72737			489.1	
PM934Q01	Societal	Employ	PM6B	15.96	(0.03)	0.59702			592.3	

[Part 3/4]

Table A.1 PISA 2012 Main Survey mathematics item classification

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Content
PM934Q02	"MATH - P2012 London Eye Q2"	Mathematics Expert Group	English	Simple Multiple Choice	Space and shape
PM936Q01	"MATH - P2012 Seats In A Theatre Q1"	Mathematics Expert Group	English	Constructed Response Manual	Change and relationships
PM936Q02	"MATH - P2012 Seats In A Theatre Q2"	Mathematics Expert Group	English	Constructed Response Expert	Change and relationships
PM939Q01	"MATH - P2012 Racing Q1"	Mathematics Expert Group	English	Simple Multiple Choice	Uncertainty and data
PM939Q02	"MATH - P2012 Racing Q2"	Mathematics Expert Group	English	Simple Multiple Choice	Uncertainty and data
PM942Q01	"MATH - P2012 Climbing Mount Fuji Q1"	ACER	English	Simple Multiple Choice	Quantity
PM942Q02	"MATH - P2012 Climbing Mount Fuji Q2"	ACER	English	Constructed Response Expert	Change and relationships
PM942Q03	"MATH - P2012 Climbing Mount Fuji Q3"	ACER	English	Constructed Response Manual	Quantity
PM943Q01	"MATH - P2012 Arches Q1"	ACER	English	Simple Multiple Choice	Change and relationships
PM943Q02	"MATH - P2012 Arches Q2"	ACER	English	Constructed Response Expert	Space and shape
PM948Q01	"MATH - P2012 Part-Time Work Q1"	ACER	English	Simple Multiple Choice	Quantity
PM948Q02	"MATH - P2012 Part-Time Work Q2"	ACER	English	Constructed Response Manual	Quantity
PM948Q03	"MATH - P2012 Part-Time Work Q3"	ACER	English	Constructed Response Expert	Quantity
PM949Q01T	"MATH - P2012 Roof Truss Design Q1"	ACER	English	Complex Multiple Choice	Space and shape
PM949Q02T	"MATH - P2012 Roof Truss Design Q2"	ACER	English	Complex Multiple Choice	Space and shape
PM949Q03	"MATH - P2012 Roof Truss Design Q3"	ACER	English	Constructed Response Expert	Space and shape
PM953Q02	"MATH - P2012 Flu Test Q2"	University of Melbourne	English	Constructed Response Expert	Uncertainty and data
PM953Q03	"MATH - P2012 Flu Test Q3"	University of Melbourne	English	Constructed Response Manual	Uncertainty and data
PM953Q04D	"MATH - P2012 Flu Test Q4"	University of Melbourne	English	Constructed Response Expert	Uncertainty and data
PM954Q01	"MATH - P2012 Medicine Doses Q1"	University of Melbourne	English	Constructed Response Manual	Change and relationships
PM954Q02	"MATH - P2012 Medicine Doses Q2"	University of Melbourne	English	Constructed Response Expert	Change and relationships
PM954Q04	"MATH - P2012 Medicine Doses Q4"	University of Melbourne	English	Constructed Response Expert	Change and relationships
PM955Q01	"MATH - P2012 Migration Q1"	University of Melbourne	English	Constructed Response Manual	Uncertainty and data
PM955Q02	"MATH - P2012 Migration Q2"	University of Melbourne	English	Constructed Response Expert	Uncertainty and data
PM955Q03	"MATH - P2012 Migration Q3"	University of Melbourne	English	Constructed Response Expert	Uncertainty and data
PM957Q01	"MATH - P2012 Helen The Cyclist Q1"	University of Melbourne	English	Simple Multiple Choice	Change and relationships
PM957Q02	"MATH - P2012 Helen The Cyclist Q2"	University of Melbourne	English	Simple Multiple Choice	Change and relationships
PM957Q03	"MATH - P2012 Helen The Cyclist Q3"	University of Melbourne	English	Constructed Response Manual	Change and relationships
PM961Q02	"MATH - P2012 Chocolate Q2"	IPN/Kassel	English	Constructed Response Expert	Change and relationships
PM961Q03	"MATH - P2012 Chocolate Q3"	IPN/Kassel	English	Simple Multiple Choice	Change and relationships
PM961Q05	"MATH - P2012 Chocolate Q5"	IPN/Kassel	English	Constructed Response Expert	Uncertainty and data
PM967Q01	"MATH - P2012 Wooden Train Set Q1"	IPN/Kassel	English	Constructed Response Manual	Space and shape
PM967Q03T	"MATH - P2012 Wooden Train Set Q3"	IPN/Kassel	English	Complex Multiple Choice	Space and shape
PM982Q01	"MATH - P2012 Employment Data Q1"	ACER	English	Constructed Response Manual	Uncertainty and data
PM982Q02	"MATH - P2012 Employment Data Q2"	ACER	English	Constructed Response Manual	Uncertainty and data
PM982Q03T	"MATH - P2012 Employment Data Q3"	ACER	English	Complex Multiple Choice	Uncertainty and data
PM982Q04	"MATH - P2012 Employment Data Q4"	ACER	English	Simple Multiple Choice	Uncertainty and data
PM985Q01	"MATH - P2012 Which Car Q1"	ACER	English	Simple Multiple Choice	Uncertainty and data
PM985Q02	"MATH - P2012 Which Car Q2"	ACER	English	Simple Multiple Choice	Quantity
PM985Q03	"MATH - P2012 Which Car Q3"	ACER	English	Constructed Response Manual	Quantity
PM991Q01	"MATH - P2012 Garage Q1"	France	English	Simple Multiple Choice	Space and shape
PM991Q02D	"MATH - P2012 Garage Q2"	France	English	Constructed Response Expert	Space and shape
PM992Q01	"MATH - P2012 Spacers Q1"	France	English	Constructed Response Manual	Space and shape
PM992Q02	"MATH - P2012 Spacers Q2"	France	English	Constructed Response Manual	Space and shape
PM992Q03	"MATH - P2012 Spacers Q3"	France	English	Constructed Response Expert	Change and relationships
PM995Q01	"MATH - P2012 Revolving Door Q1"	ILS	English	Constructed Response Manual	Space and shape
PM995Q02	"MATH - P2012 Revolving Door Q2"	ILS	English	Constructed Response Expert	Space and shape
PM995Q03	"MATH - P2012 Revolving Door Q3"	ILS	English	Simple Multiple Choice	Quantity
PM998Q02	"MATH - P2012 Bike Rental Q2"	Israel	English	Constructed Response Manual	Change and relationships
PM998Q04T	"MATH - P2012 Bike Rental Q4"	Israel	English	Complex Multiple Choice	Change and relationships

[Part 4/4]

Table A.1 PISA 2012 Main Survey mathematics item classification

Unit Item Code	Context	Process	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
						Delta	Tau(1)	Tau(2)	1	2
PM934Q02	Societal	Formulate	PM6B	43.60	(0.05)	-0.83074			481.0	
PM936Q01	Occupational	Employ	PM7B	28.95	(0.05)	-0.64450			495.5	
PM936Q02	Occupational	Formulate	PM7B	30.36	(0.04)	-0.47460			508.8	
PM939Q01	Societal	Interpret	PM7B	45.71	(0.04)	-0.94709			472.0	
PM939Q02	Societal	Interpret	PM7B	38.14	(0.05)	-0.38637			515.6	
PM942Q01	Societal	Formulate	PM6B	46.93	(0.05)	-1.04973			464.0	
PM942Q02	Societal	Formulate	PM6B	14.25	(0.03)	1.23043			641.6	
PM942Q03	Societal	Employ	PM6B	11.58	(0.03)	0.70467	1.42884	-1.42884	591.3	610.0
PM943Q01	Occupational	Formulate	PM7A	50.02	(0.24)	-0.18842			531.1	
PM943Q02	Occupational	Formulate	PM7A	5.29	(0.11)	3.07526			785.3	
PM948Q01	Occupational	Interpret	PM7B	85.91	(0.03)	-2.79957			327.7	
PM948Q02	Occupational	Employ	PM7B	64.13	(0.05)	-1.73697			410.5	
PM948Q03	Occupational	Employ	PM7B	8.84	(0.02)	1.76132			683.0	
PM949Q01T	Occupational	Employ	PM5,PMUH	67.51	(0.23)	-1.00898			467.2	
PM949Q02T	Occupational	Employ	PM5,PMUH	31.74	(0.23)	0.62188			594.2	
PM949Q03	Occupational	Formulate	PM5,PMUH	32.55	(0.23)	0.45241	2.35997	-2.35997	577.3	584.7
PM953Q02	Scientific	Interpret	PM7A	49.76	(0.24)	0.00698			546.3	
PM953Q03	Scientific	Formulate	PM7A	51.80	(0.25)	-0.12631			536.0	
PM953Q04D	Scientific	Formulate	PM7A	18.21	(0.18)	1.46739	0.83011	-0.83011	643.2	676.9
PM954Q01	Scientific	Employ	PM7A	65.40	(0.23)	-1.00422			467.6	
PM954Q02	Scientific	Employ	PM7A	33.56	(0.23)	0.88774			614.9	
PM954Q04	Scientific	Employ	PM7A	26.35	(0.21)	1.26195			644.1	
PM955Q01	Societal	Interpret	PM5	72.12	(0.21)	-1.21749			450.9	
PM955Q02	Societal	Interpret	PM5	34.21	(0.23)	0.66748			597.7	
PM955Q03	Societal	Employ	PM5	11.98	(0.15)	1.81095	0.98867	-0.98867	672.4	701.3
PM957Q01	Personal	Employ	PM6B	52.91	(0.05)	-1.35141			440.5	
PM957Q02	Personal	Employ	PM6B	36.86	(0.04)	-0.45100			510.6	
PM957Q03	Personal	Employ	PM6B	5.75	(0.02)	1.93528			696.6	
PM961Q02	Occupational	Employ	PM7B	4.21	(0.02)	2.20028			717.1	
PM961Q03	Scientific	Employ	PM7B	44.68	(0.05)	-0.81403			482.4	
PM961Q05	Occupational	Interpret	PM7B	42.76	(0.04)	-0.75782	-0.00894	0.00894	449.0	524.6
PM967Q01	Personal	Employ	PM7B	30.21	(0.05)	0.08827			552.6	
PM967Q03T	Personal	Formulate	PM7B	7.03	(0.03)	1.66062			675.2	
PM982Q01	Societal	Employ	PM4	87.30	(0.16)	-2.48212			352.4	
PM982Q02	Societal	Employ	PM4	30.73	(0.21)	0.81899			609.6	
PM982Q03T	Societal	Interpret	PM4	64.95	(0.23)	-1.00060			467.8	
PM982Q04	Societal	Formulate	PM4	51.45	(0.24)	-0.17651			532.0	
PM985Q01	Personal	Interpret	PM6B	81.14	(0.03)	-2.79792			327.8	
PM985Q02	Personal	Employ	PM6B	37.48	(0.05)	-0.70435			490.9	
PM985Q03	Personal	Employ	PM6B	25.56	(0.04)	0.08777			552.6	
PM991Q01	Occupational	Interpret	PM6B	65.14	(0.05)	-1.62018			419.6	
PM991Q02D	Occupational	Employ	PM6B	2.66	(0.01)	1.66230	1.16523	-1.16523	663.2	687.3
PM992Q01	Occupational	Formulate	PM4,PMUH	77.60	(0.20)	-1.75452			409.1	
PM992Q02	Occupational	Formulate	PM4,PMUH	18.25	(0.19)	1.62983			672.7	
PM992Q03	Occupational	Formulate	PM4,PMUH	8.11	(0.14)	2.63540			751.1	
PM995Q01	Scientific	Employ	PM6A	57.67	(0.25)	-0.42930			512.3	
PM995Q02	Scientific	Formulate	PM6A	3.47	(0.08)	3.78162			840.3	
PM995Q03	Scientific	Formulate	PM6A	46.42	(0.24)	0.19919			561.3	
PM998Q02	Personal	Interpret	PM5	71.57	(0.21)	-1.08005			461.6	
PM998Q04T	Personal	Employ	PM5	40.44	(0.23)	0.29426			568.7	

[Part 1/2]

Table A.2 **PISA 2012 Main Survey reading item classification**

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Situation
PR220Q01	“READ - P2000 South Pole Q1”	France	French	Constructed Response Expert	Educational
PR220Q02B	“READ - P2000 South Pole Q2B”	France	French	Simple Multiple Choice	Educational
PR220Q04	“READ - P2000 South Pole Q4”	France	French	Simple Multiple Choice	Educational
PR404Q03	“READ - P2009 Sleep Q3”	ILS	Norwegian	Simple Multiple Choice	Public
PR404Q06	“READ - P2009 Sleep Q6”	ILS	Norwegian	Simple Multiple Choice	Public
PR404Q07T	“READ - P2009 Sleep Q7”	ILS	Norwegian	Complex Multiple Choice	Public
PR404Q10A	“READ - P2009 Sleep Q10A”	ILS	Norwegian	Constructed Response Expert	Public
PR404Q10B	“READ - P2009 Sleep Q10B”	ILS	Norwegian	Constructed Response Expert	Public
PR406Q01	“READ - P2009 Kokeshi Dolls Q1”	NIER	Japanese	Constructed Response Expert	Personal
PR406Q02	“READ - P2009 Kokeshi Dolls Q2”	NIER	Japanese	Constructed Response Expert	Personal
PR406Q05	“READ - P2009 Kokeshi Dolls Q5”	NIER	Japanese	Constructed Response Expert	Personal
PR412Q01	“READ - P2009 World Languages Q1”	ACER	English	Simple Multiple Choice	Educational
PR412Q05	“READ - P2009 World Languages Q5”	ACER	English	Simple Multiple Choice	Educational
PR412Q06T	“READ - P2009 World Languages Q6”	ACER	English	Complex Multiple Choice	Educational
PR412Q08	“READ - P2009 World Languages Q8”	ACER	English	Constructed Response Expert	Educational
PR420Q02	“READ - P2009 Childrens Futures Q2”	NIER	Japanese	Constructed Response Manual	Educational
PR420Q06	“READ - P2009 Childrens Futures Q6”	NIER	Japanese	Constructed Response Expert	Educational
PR420Q09	“READ - P2009 Childrens Futures Q9”	NIER	Japanese	Constructed Response Manual	Educational
PR420Q10	“READ - P2009 Childrens Futures Q10”	NIER	Japanese	Constructed Response Expert	Educational
PR424Q02T	“READ - P2009 Fair Trade Q2”	aSPe	French	Complex Multiple Choice	Educational
PR424Q03	“READ - P2009 Fair Trade Q3”	aSPe	French	Simple Multiple Choice	Educational
PR424Q07	“READ - P2009 Fair Trade Q7”	aSPe	French	Simple Multiple Choice	Educational
PR432Q01	“READ - P2009 About a book Q1”	DIPF	German	Constructed Response Manual	Personal
PR432Q05	“READ - P2009 About a book Q5”	DIPF	German	Constructed Response Expert	Personal
PR432Q06T	“READ - P2009 About a book Q6”	DIPF	German	Complex Multiple Choice	Personal
PR437Q01	“READ - P2009 Narcissus Q1”	Sweden	Portuguese	Simple Multiple Choice	Personal
PR437Q06	“READ - P2009 Narcissus Q6”	Sweden	Portuguese	Simple Multiple Choice	Personal
PR437Q07	“READ - P2009 Narcissus Q7”	Sweden	Portuguese	Constructed Response Expert	Personal
PR446Q03	“READ - P2009 Job Vacancy Q3”	ACER	English	Constructed Response Manual	Occupational
PR446Q06	“READ - P2009 Job Vacancy Q6”	ACER	English	Constructed Response Expert	Occupational
PR453Q01	“READ - P2009 Find Summer Job Q1”	Finland	Finnish	Simple Multiple Choice	Occupational
PR453Q04	“READ - P2009 Find Summer Job Q4”	Finland	Finnish	Constructed Response Expert	Occupational
PR453Q05T	“READ - P2009 Find Summer Job Q5”	Finland	Finnish	Complex Multiple Choice	Occupational
PR453Q06	“READ - P2009 Find Summer Job Q6”	Finland	Finnish	Constructed Response Expert	Occupational
PR455Q02	“READ - P2009 Chocolate and Health Q2”	New Zealand	English	Constructed Response Expert	Personal
PR455Q03	“READ - P2009 Chocolate and Health Q3”	New Zealand	English	Constructed Response Manual	Personal
PR455Q04	“READ - P2009 Chocolate and Health Q4”	New Zealand	English	Simple Multiple Choice	Personal
PR455Q05T	“READ - P2009 Chocolate and Health Q5”	New Zealand	English	Complex Multiple Choice	Personal
PR456Q01	“READ - P2009 Biscuits Q1”	Serbia	English	Simple Multiple Choice	Personal
PR456Q02	“READ - P2009 Biscuits Q2”	Serbia	English	Constructed Response Expert	Personal
PR456Q06	“READ - P2009 Biscuits Q6”	Serbia	English	Constructed Response Expert	Personal
PR466Q02	“READ - P2009 Work Right Q2”	aSPe	French	Constructed Response Expert	Occupational
PR466Q03T	“READ - P2009 Work Right Q3”	aSPe	French	Complex Multiple Choice	Occupational
PR466Q06	“READ - P2009 Work Right Q6”	aSPe	French	Constructed Response Manual	Occupational

[Part 2/2]

Table A.2 **PISA 2012 Main Survey reading item classification**

Unit Item Code	Scale		Aspect	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
	Text Format	Text Type					Delta	Tau(1)	Tau(2)	1	2
PR220Q01	Mixed	Exposition	Access and retrieve	PR3	38.53	(0.24)	1.40904			694.1	
PR220Q02B	Mixed	Exposition	Integrate and interpret	PR3	60.99	(0.24)	0.07003			586.7	
PR220Q04	Continuous	Exposition	Integrate and interpret	PR3	57.67	(0.24)	0.27950			603.4	
PR404Q03	Continuous	Exposition	Integrate and interpret	PR1	74.00	(0.22)	-0.87766			510.6	
PR404Q06	Non-continuous	Exposition	Integrate and interpret	PR1	50.08	(0.23)	0.58923			628.3	
PR404Q07T	Non-continuous	Exposition	Integrate and interpret	PR1	36.46	(0.22)	1.40667			694.0	
PR404Q10A	Non-continuous	Exposition	Reflect and evaluate	PR1	46.58	(0.24)	0.85188			649.4	
PR404Q10B	Non-continuous	Exposition	Reflect and evaluate	PR1	37.39	(0.23)	1.21553			678.6	
PR406Q01	Continuous	Narration	Integrate and interpret	PR1	65.61	(0.24)	-0.42613			546.9	
PR406Q02	Continuous	Narration	Integrate and interpret	PR1	34.72	(0.22)	1.11002			670.1	
PR406Q05	Continuous	Narration	Integrate and interpret	PR1	74.11	(0.21)	-0.74389			521.3	
PR412Q01	Non-continuous	Exposition	Access and retrieve	PR2	85.36	(0.16)	-1.64252			449.2	
PR412Q05	Continuous	Exposition	Integrate and interpret	PR2	58.34	(0.23)	0.09273			588.5	
PR412Q06T	Continuous	Exposition	Integrate and interpret	PR2	38.98	(0.22)	0.89309			652.7	
PR412Q08	Mixed	Exposition	Integrate and interpret	PR2	36.84	(0.23)	1.22161			679.1	
PR420Q02	Non-continuous	Exposition	Access and retrieve	PR2	84.57	(0.17)	-1.41503			467.5	
PR420Q06	Non-continuous	Exposition	Reflect and evaluate	PR2	45.28	(0.22)	0.71132			638.1	
PR420Q09	Non-continuous	Exposition	Access and retrieve	PR2	76.53	(0.19)	-1.07532			494.8	
PR420Q10	Non-continuous	Exposition	Integrate and interpret	PR2	73.48	(0.21)	-0.36037	2.01069	-2.01069	546.8	557.5
PR424Q02T	Non-continuous	Argumentation	Integrate and interpret	PR1	44.17	(0.24)	0.85301			649.5	
PR424Q03	Non-continuous	Argumentation	Reflect and evaluate	PR1	65.46	(0.22)	-0.30577			556.5	
PR424Q07	Continuous	Argumentation	Reflect and evaluate	PR1	76.25	(0.20)	-1.03249			498.2	
PR432Q01	Continuous	Argumentation	Integrate and interpret	PR3	85.88	(0.16)	-1.50920			459.9	
PR432Q05	Multiple	Argumentation	Reflect and evaluate	PR3	75.72	(0.21)	-0.52250			539.2	
PR432Q06T	Continuous	Argumentation	Integrate and interpret	PR3	14.99	(0.16)	2.88148			812.4	
PR437Q01	Continuous	Narration	Integrate and interpret	PR2	52.14	(0.23)	0.35405			609.5	
PR437Q06	Continuous	Narration	Integrate and interpret	PR2	53.72	(0.22)	0.29842			605.0	
PR437Q07	Continuous	Narration	Integrate and interpret	PR2	14.84	(0.16)	2.38872			772.8	
PR446Q03	Non-continuous	Description	Access and retrieve	PR3	92.93	(0.13)	-2.32901			394.1	
PR446Q06	Non-continuous	Description	Reflect and evaluate	PR3	77.56	(0.20)	-0.82754			514.7	
PR453Q01	Continuous	Instruction	Integrate and interpret	PR2	82.07	(0.18)	-1.14369			489.2	
PR453Q04	Continuous	Instruction	Reflect and evaluate	PR2	62.36	(0.22)	-0.06473			575.9	
PR453Q05T	Continuous	Instruction	Access and retrieve	PR2	63.34	(0.23)	0.02273			582.8	
PR453Q06	Continuous	Instruction	Reflect and evaluate	PR2	72.03	(0.21)	-0.37494			551.0	
PR455Q02	Continuous	Description	Reflect and evaluate	PR1,PRUH	38.52	(0.23)	1.24380			680.9	
PR455Q03	Continuous	Description	Access and retrieve	PR1,PRUH	76.07	(0.21)	-1.17540			486.7	
PR455Q04	Continuous	Description	Integrate and interpret	PR1,PRUH	63.26	(0.23)	-0.21940			563.4	
PR455Q05T	Continuous	Description	Integrate and interpret	PR1,PRUH	26.08	(0.22)	1.98631			740.4	
PR456Q01	Continuous	Narration	Access and retrieve	PR3,PRUH	96.22	(0.09)	-3.33252			313.6	
PR456Q02	Continuous	Narration	Integrate and interpret	PR3,PRUH	83.50	(0.17)	-1.33872			473.6	
PR456Q06	Continuous	Narration	Integrate and interpret	PR3,PRUH	83.79	(0.18)	-1.47870			462.4	
PR466Q02	Continuous	Argumentation	Access and retrieve	PR3	44.35	(0.24)	0.80882			646.0	
PR466Q03T	Mixed	Argumentation	Integrate and interpret	PR3	16.98	(0.18)	2.63235			792.3	
PR466Q06	Continuous	Argumentation	Access and retrieve	PR3	81.43	(0.19)	-1.12481			490.8	

[Part 1/2]

Table A.3 **PISA 2012 Main Survey science item classification**

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Application Area	Item Focus
PS131Q02D	"SCIE - P2000 Good Vibrations Q2"	ACER	English	Constructed Response Expert	Health	Social
PS131Q04D	"SCIE - P2006 Good Vibrations Q4"	ACER	English	Constructed Response Expert	Health	Social
PS256Q01	"SCIE - P2000 Spoons Q1"	TIMSS	English	Simple Multiple Choice	Frontiers	Personal
PS269Q01	"SCIE - P2000 Earths Temperature Q1"	CITO	Dutch	Constructed Response Expert	Hazards	Global
PS269Q03D	"SCIE - P2000 Earths Temperature Q3"	CITO	Dutch	Constructed Response Expert	Hazards	Global
PS269Q04T	"SCIE - P2000 Earths Temperature Q4"	CITO	Dutch	Complex Multiple Choice	Natural resources	Global
PS326Q01	"SCIE - P2003 Milk Q1"	CITO	Dutch	Constructed Response Expert	Health	Personal
PS326Q02	"SCIE - P2003 Milk Q2"	CITO	Dutch	Constructed Response Expert	Health	Personal
PS326Q03	"SCIE - P2003 Milk Q3"	CITO	Dutch	Simple Multiple Choice	Health	Personal
PS326Q04T	"SCIE - P2003 Milk Q4"	CITO	Dutch	Complex Multiple Choice	Health	Personal
PS408Q01	"SCIE - P2006 Wild Oat Grass Q1"	ILS	Norwegian	Simple Multiple Choice	Natural resources	Social
PS408Q03	"SCIE - P2006 Wild Oat Grass Q3"	ILS	Norwegian	Constructed Response Expert	Natural resources	Social
PS408Q04T	"SCIE - P2006 Wild Oat Grass Q4"	ILS	Norwegian	Complex Multiple Choice	Natural resources	Social
PS408Q05	"SCIE - P2006 Wild Oat Grass Q5"	ILS	Norwegian	Simple Multiple Choice	Natural resources	Social
PS413Q04T	"SCIE - P2006 Plastic Age Q4"	IPN	German	Complex Multiple Choice	Frontiers	Social
PS413Q05	"SCIE - P2006 Plastic Age Q5"	IPN	German	Simple Multiple Choice	Frontiers	Social
PS413Q06	"SCIE - P2006 Plastic Age Q6"	IPN	German	Constructed Response Manual	Frontiers	Personal
PS415Q02	"SCIE - P2006 Solar Panels Q2"	NIER	Japanese	Simple Multiple Choice	Natural resources	Global
PS415Q07T	"SCIE - P2006 Solar Panels Q7"	ACER	English	Complex Multiple Choice	Natural resources	Personal
PS415Q08T	"SCIE - P2006 Solar Panels Q8"	ACER	English	Complex Multiple Choice	Natural resources	Global
PS425Q02	"SCIE - P2006 Penguin Island Q2"	ACER	English	Simple Multiple Choice	Environment	Social
PS425Q03	"SCIE - P2006 Penguin Island Q3"	ACER	English	Constructed Response Expert	Environment	Social
PS425Q04	"SCIE - P2006 Penguin Island Q4"	ACER	English	Constructed Response Expert	Environment	Social
PS425Q05	"SCIE - P2006 Penguin Island Q5"	ACER	English	Simple Multiple Choice	Environment	Social
PS428Q01	"SCIE - P2006 Bacteria in Milk Q1"	IPN	German	Simple Multiple Choice	Natural resources	Social
PS428Q03	"SCIE - P2006 Bacteria in Milk Q3"	IPN	German	Simple Multiple Choice	Natural resources	Social
PS428Q05	"SCIE - P2006 Bacteria in Milk Q5"	IPN	German	Constructed Response Expert	Natural resources	Social
PS438Q01T	"SCIE - P2006 Green Parks Q1"	ACER	English	Complex Multiple Choice	Environment	Social
PS438Q02	"SCIE - P2006 Green Parks Q2"	ACER	English	Simple Multiple Choice	Environment	Social
PS438Q03D	"SCIE - P2006 Green Parks Q3"	ACER	English	Constructed Response Expert	Environment	Social
PS465Q01	"SCIE - P2006 Different Climates Q1"	ILS	Norwegian	Constructed Response Expert	Environment	Global
PS465Q02	"SCIE - P2006 Different Climates Q2"	ILS	Norwegian	Simple Multiple Choice	Environment	Global
PS465Q04	"SCIE - P2006 Different Climates Q4"	ILS	Norwegian	Simple Multiple Choice	Environment	Global
PS466Q01T	"SCIE - P2006 Forest Fires Q1"	ILS	Norwegian	Complex Multiple Choice	Hazards	Social
PS466Q05	"SCIE - P2006 Forest Fires Q5"	ILS	Norwegian	Simple Multiple Choice	Hazards	Social
PS466Q07T	"SCIE - P2006 Forest Fires Q7"	ILS	Norwegian	Complex Multiple Choice	Hazards	Social
PS478Q01	"SCIE - P2006 Antibiotics Q1"	France	French	Simple Multiple Choice	Health	Personal
PS478Q02T	"SCIE - P2006 Antibiotics Q2"	France	French	Complex Multiple Choice	Health	Personal
PS478Q03T	"SCIE - P2006 Antibiotics Q3"	France	French	Complex Multiple Choice	Health	Personal
PS498Q02T	"SCIE - P2006 Experimental Digestion Q2"	France	French	Complex Multiple Choice	Other	Social
PS498Q03	"SCIE - P2006 Experimental Digestion Q3"	France	French	Simple Multiple Choice	Other	Social
PS498Q04	"SCIE - P2006 Experimental Digestion Q4"	France	French	Constructed Response Expert	Other	Social
PS514Q02	"SCIE - P2006 Development and Disaster Q2"	NIER	Japanese	Constructed Response Expert	Hazards	Social
PS514Q03	"SCIE - P2006 Development and Disaster Q3"	NIER	Japanese	Constructed Response Expert	Hazards	Social
PS514Q04	"SCIE - P2006 Development and Disaster Q4"	NIER	Japanese	Constructed Response Manual	Hazards	Social
PS519Q01	"SCIE - P2006 Airbags Q1"	France	French	Constructed Response Expert	Frontiers	Social
PS519Q02T	"SCIE - P2006 Airbags Q2"	France	French	Complex Multiple Choice	Frontiers	Social
PS519Q03	"SCIE - P2006 Airbags Q3"	France	French	Constructed Response Expert	Frontiers	Social
PS521Q02	"SCIE - P2006 Cooking Outdoors Q2"	ACER	English	Simple Multiple Choice	Frontiers	Personal
PS521Q06	"SCIE - P2006 Cooking Outdoors Q6"	ACER	English	Simple Multiple Choice	Frontiers	Personal
PS527Q01T	"SCIE - P2006 Extinction of the Dinosaurs Q1"	Korea	Korean	Complex Multiple Choice	Frontiers	Global
PS527Q03T	"SCIE - P2006 Extinction of the Dinosaurs Q3"	Korea	Korean	Complex Multiple Choice	Frontiers	Global
PS527Q04T	"SCIE - P2006 Extinction of the Dinosaurs Q4"	Korea	Korean	Complex Multiple Choice	Frontiers	Global

[Part 2/2]

Table A.3 PISA 2012 Main Survey science item classification

Unit Item Code	Competency	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
					Delta	Tau(1)	Tau(2)	1	2
PS131Q02D	Using scientific evidence	PS3	49.41	(0.24)	0.34149			562.2	
PS131Q04D	Identifying scientific issues	PS3	28.02	(0.21)	1.31699			653.1	
PS256Q01	Explaining phenomena scientifically	PS2,PSUH	88.43	(0.15)	-2.32693			313.3	
PS269Q01	Explaining phenomena scientifically	PS1	56.94	(0.23)	-0.21324			510.4	
PS269Q03D	Explaining phenomena scientifically	PS1	43.44	(0.24)	0.47905			575.0	
PS269Q04T	Explaining phenomena scientifically	PS1	35.19	(0.22)	0.90179			614.4	
PS326Q01	Using scientific evidence	PS2	58.15	(0.23)	-0.28883			503.4	
PS326Q02	Using scientific evidence	PS2	63.64	(0.24)	-0.33764			498.8	
PS326Q03	Using scientific evidence	PS2	60.97	(0.23)	-0.18199			513.3	
PS326Q04T	Explaining phenomena scientifically	PS2	26.60	(0.21)	1.44852			665.3	
PS408Q01	Explaining phenomena scientifically	PS1	61.48	(0.23)	-0.43304			489.9	
PS408Q03	Explaining phenomena scientifically	PS1	28.67	(0.20)	1.19253			641.5	
PS408Q04T	Explaining phenomena scientifically	PS1	52.79	(0.22)	0.00221			530.5	
PS408Q05	Identifying scientific issues	PS1	42.19	(0.22)	0.47893			575.0	
PS413Q04T	Using scientific evidence	PS2	43.75	(0.23)	0.53138			579.8	
PS413Q05	Using scientific evidence	PS2	68.27	(0.22)	-0.86374			449.7	
PS413Q06	Explaining phenomena scientifically	PS2	40.31	(0.24)	0.70571			596.1	
PS415Q02	Explaining phenomena scientifically	PS3	77.95	(0.20)	-1.20447			418.0	
PS415Q07T	Identifying scientific issues	PS3	74.36	(0.21)	-0.89484			446.8	
PS415Q08T	Identifying scientific issues	PS3	59.95	(0.24)	-0.18820			512.8	
PS425Q02	Using scientific evidence	PS2	49.04	(0.25)	0.19136			548.1	
PS425Q03	Explaining phenomena scientifically	PS2	43.07	(0.23)	0.41430			568.9	
PS425Q04	Using scientific evidence	PS2	28.92	(0.22)	1.32550			653.9	
PS425Q05	Identifying scientific issues	PS2	68.38	(0.21)	-0.66175			468.6	
PS428Q01	Using scientific evidence	PS3,PSUH	61.46	(0.23)	-0.28896			503.4	
PS428Q03	Using scientific evidence	PS3,PSUH	74.31	(0.21)	-0.92135			444.4	
PS428Q05	Explaining phenomena scientifically	PS3,PSUH	46.37	(0.24)	0.41594			569.1	
PS438Q01T	Identifying scientific issues	PS3	82.88	(0.18)	-1.37034			402.6	
PS438Q02	Identifying scientific issues	PS3	66.66	(0.23)	-0.53754			480.1	
PS438Q03D	Identifying scientific issues	PS3	38.55	(0.24)	0.95762			619.6	
PS465Q01	Using scientific evidence	PS3	47.85	(0.21)	0.40387	0.03760	-0.03760	524.6	611.3
PS465Q02	Explaining phenomena scientifically	PS3	59.65	(0.23)	-0.18291			513.2	
PS465Q04	Explaining phenomena scientifically	PS3	36.90	(0.22)	0.77485			602.6	
PS466Q01T	Identifying scientific issues	PS1,PSUH	73.61	(0.21)	-0.87507			448.7	
PS466Q05	Using scientific evidence	PS1,PSUH	53.44	(0.23)	-0.05052			525.5	
PS466Q07T	Identifying scientific issues	PS1,PSUH	70.08	(0.21)	-0.79338			456.4	
PS478Q01	Explaining phenomena scientifically	PS2	44.71	(0.22)	0.41446			568.9	
PS478Q02T	Using scientific evidence	PS2	55.83	(0.24)	0.01756			532.0	
PS478Q03T	Explaining phenomena scientifically	PS2	70.35	(0.20)	-0.78431			457.2	
PS498Q02T	Identifying scientific issues	PS2	45.07	(0.23)	0.30482			558.7	
PS498Q03	Identifying scientific issues	PS2	38.64	(0.22)	0.56938			583.4	
PS498Q04	Using scientific evidence	PS2	63.91	(0.23)	-0.30115	0.97189	-0.97189	484.7	519.8
PS514Q02	Using scientific evidence	PS3	84.82	(0.18)	-1.52005			388.6	
PS514Q03	Explaining phenomena scientifically	PS3	46.42	(0.23)	0.44292			571.6	
PS514Q04	Using scientific evidence	PS3	57.80	(0.24)	-0.14639			516.7	
PS519Q01	Using scientific evidence	PS1	37.73	(0.20)	0.59545	0.28465	-0.28465	551.6	620.1
PS519Q02T	Explaining phenomena scientifically	PS1	55.21	(0.23)	-0.09256			521.6	
PS519Q03	Identifying scientific issues	PS1	25.57	(0.20)	1.34784			656.0	
PS521Q02	Explaining phenomena scientifically	PS1	53.30	(0.22)	-0.01855			528.5	
PS521Q06	Explaining phenomena scientifically	PS1	88.78	(0.15)	-2.12856			331.8	
PS527Q01T	Using scientific evidence	PS1	17.46	(0.17)	2.04093			720.6	
PS527Q03T	Explaining phenomena scientifically	PS1	57.62	(0.23)	-0.08024			522.8	
PS527Q04T	Explaining phenomena scientifically	PS1	54.87	(0.23)	0.07115			536.9	

[Part 1/2]

Table A.4 PISA 2012 Main Survey digital reading item classification

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Situation	Text Format	Text Type
CR002Q01	Seraing	aSPe	French	Simple Multiple Choice	Public	Non-continuous	Description
CR002Q03	Seraing	aSPe	French	Simple Multiple Choice	Public	Multiple	Description
CR002Q05	Seraing	aSPe	French	Constructed Response Expert	Personal	Multiple	Transaction
CR011Q01A	Cinema	ACER	English	Selected Response Variations	Personal	Multiple	Transaction
CR011Q01B	Cinema	ACER	English	Constructed Response Auto-coded	Personal	Multiple	Transaction
CR013Q01	Sports Club	DIPF	German	Simple Multiple Choice	Personal	Multiple	Transaction
CR013Q04	Sports Club	DIPF	German	Simple Multiple Choice	Personal	Multiple	Argumentation
CR013Q07	Sports Club	DIPF	German	Constructed Response Expert	Personal	Multiple	Argumentation
CR014Q01	Hay Fever	DIPF	German	Constructed Response Expert	Public	Non-continuous	Description
CR014Q06	Hay Fever	DIPF	German	Simple Multiple Choice	Public	Multiple	Exposition
CR014Q07	Hay Fever	DIPF	German	Simple Multiple Choice	Public	Multiple	Description
CR014Q11	Hay Fever	DIPF	German	Simple Multiple Choice	Public	Multiple	Exposition
CR017Q01	Language Learning	DIPF	German	Simple Multiple Choice	Educational	Continuous	Argumentation
CR017Q04	Language Learning	DIPF	German	Simple Multiple Choice	Educational	Multiple	Argumentation
CR017Q07	Language Learning	DIPF	German	Constructed Response Expert	Educational	Multiple	Argumentation
CR021Q01	Counterfeiting	Canada	English/French	Simple Multiple Choice	Public	Mixed	Exposition
CR021Q04	Counterfeiting	Canada	English/French	Simple Multiple Choice	Public	Multiple	Exposition
CR021Q05	Counterfeiting	Canada	English/French	Simple Multiple Choice	Public	Multiple	Exposition
CR021Q08	Counterfeiting	Canada	English/French	Constructed Response Expert	Public	Multiple	Exposition

[Part 2/2]

Table A.4 PISA 2012 Main Survey digital reading item classification

Unit Item Code	Aspect	Environment	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)				Thresholds (RP=0.62) PISA scale		
						Delta	Tau(1)	Tau(2)	Tau(3)	1	2	3
CR002Q01	Access and retrieve	Authored	CR2	93.85	(0.15)	-2.78159				247.8		
CR002Q03	Access and retrieve	Authored	CR2	81.06	(0.25)	-1.24934				381.9		
CR002Q05	Complex	Mixed	CR2	45.48	(0.30)	0.76767	1.36902	-1.36902		547.4	569.6	
CR011Q01A	Complex	Message-based	CR1	71.42	(0.29)	-0.52991				444.9		
CR011Q01B	Complex	Message-based	CR1	69.12	(0.26)	-0.17018	0.33163	-0.33163		445.6	507.2	
CR013Q01	Integrate and interpret	Message-based	CR2	67.71	(0.29)	-0.31371				463.9		
CR013Q04	Access and retrieve	Authored	CR2	65.21	(0.29)	-0.12498				480.3		
CR013Q07	Complex	Authored	CR2	45.93	(0.29)	0.71723	0.75550	-0.75550		533.6	574.4	
CR014Q01	Reflect and evaluate	Authored	CR1	45.78	(0.25)	0.77642	-0.52319	0.52319		492.1	626.3	
CR014Q06	Access and retrieve	Authored	CR1	48.89	(0.31)	0.62422				545.9		
CR014Q07	Integrate and interpret	Authored	CR1	60.93	(0.30)	-0.00166				491.1		
CR014Q11	Integrate and interpret	Authored	CR1	42.50	(0.29)	1.00942				579.7		
CR017Q01	Integrate and interpret	Authored	CR2	52.21	(0.28)	0.67218				550.1		
CR017Q04	Access and retrieve	Message-based	CR2	91.06	(0.18)	-2.05505				311.4		
CR017Q07	Reflect and evaluate	Message-based	CR2	47.20	(0.25)	0.61828	-0.26649	0.26649		491.7	599.1	
CR021Q01	Integrate and interpret	Authored	CR1	57.77	(0.31)	0.19695				508.5		
CR021Q04	Reflect and evaluate	Authored	CR1	70.29	(0.29)	-0.48182				449.2		
CR021Q05	Integrate and interpret	Authored	CR1	63.11	(0.30)	-0.03197				488.5		
CR021Q08	Complex	Authored	CR1	12.54	(0.14)	2.35784	0.39213	-1.13453	0.74200	637.6	669.7	778.8

[Part 1/2]

Table A.5 **PISA 2012 Main Survey computer-based mathematics item classification**

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Content	Context
CM002Q01	Cube building	ACER	English	Constructed Response Auto-coded	Space and shape	Personal
CM002Q02	Cube building	ACER	English	Constructed Response Auto-coded	Space and shape	Personal
CM004Q01T	Toast	University of Melbourne	English	Complex Multiple Choice	Space and shape	Personal
CM004Q02T	Toast	University of Melbourne	English	Complex Multiple Choice	Space and shape	Personal
CM004Q03	Toast	University of Melbourne	English	Constructed Response Auto-coded	Space and shape	Personal
CM005Q01	Fit the curve	ACER	English	Constructed Response Auto-coded	Change and relationships	Scientific
CM005Q02	Fit the curve	ACER	English	Simple Multiple Choice	Change and relationships	Scientific
CM005Q03	Fit the curve	ACER	English	Simple Multiple Choice	Change and relationships	Scientific
CM005Q04	Fit the curve	ACER	English	Constructed Response Auto-coded	Change and relationships	Scientific
CM006Q01	Buying online	ACER	English	Constructed Response Auto-coded	Quantity	Personal
CM006Q02	Buying online	ACER	English	Constructed Response Auto-coded	Quantity	Personal
CM006Q03	Buying online	ACER	English	Constructed Response Auto-coded	Quantity	Personal
CM008Q01	Measure	ACER	English	Constructed Response Auto-coded	Space and shape	Societal
CM008Q02	Measure	ACER	English	Constructed Response Auto-coded	Quantity	Societal
CM011Q01T	Train or car	IPN/Kassel	English	Selected Response Variations	Change and relationships	Societal
CM011Q02	Train or car	IPN/Kassel	English	Constructed Response Auto-coded	Change and relationships	Societal
CM011Q03	Train or car	IPN/Kassel	English	Constructed Response Auto-coded	Change and relationships	Societal
CM014Q01	Cotton	ACER	English	Constructed Response Auto-coded	Uncertainty and data	Occupational
CM014Q02	Cotton	ACER	English	Constructed Response Auto-coded	Uncertainty and data	Occupational
CM014Q03T	Cotton	ACER	English	Selected Response Variations	Uncertainty and data	Occupational
CM015Q01	CD production	ACER	English	Simple Multiple Choice	Quantity	Occupational
CM015Q02D	CD production	ACER	English	Constructed Response Auto-coded	Change and relationships	Occupational
CM015Q03D	CD production	ACER	English	Constructed Response Expert	Change and relationships	Occupational
CM016Q01	Shelving library books	ACER	English	Constructed Response Auto-coded	Quantity	Occupational
CM020Q01	Star points	IPN/Kassel	English	Constructed Response Auto-coded	Space and shape	Scientific
CM020Q02	Star points	IPN/Kassel	English	Simple Multiple Choice	Space and shape	Scientific
CM020Q03	Star points	IPN/Kassel	English	Constructed Response Auto-coded	Space and shape	Scientific
CM020Q04	Star points	IPN/Kassel	English	Simple Multiple Choice	Space and shape	Scientific
CM025Q01	Common injuries	ACER	English	Constructed Response Auto-coded	Uncertainty and data	Occupational
CM025Q02T	Common injuries	ACER	English	Selected Response Variations	Uncertainty and data	Occupational
CM028Q03	Shopping calculations	France	French	Constructed Response Expert	Quantity	Personal
CM035Q01	GPS navigation	ACER	English	Simple Multiple Choice	Space and shape	Personal
CM035Q02	GPS navigation	ACER	English	Simple Multiple Choice	Quantity	Personal
CM035Q03	GPS navigation	ACER	English	Simple Multiple Choice	Space and shape	Personal
CM035Q04	GPS navigation	ACER	English	Constructed Response Auto-coded	Quantity	Personal
CM036Q01T	Cherry blossoms	NIER	English	Complex Multiple Choice	Uncertainty and data	Societal
CM036Q02	Cherry blossoms	NIER	English	Constructed Response Auto-coded	Change and relationships	Societal
CM036Q03	Cherry blossoms	NIER	English	Constructed Response Auto-coded	Change and relationships	Societal
CM038Q03T	Body mass index	Macau/Serbia	English	Complex Multiple Choice	Uncertainty and data	Societal
CM038Q05	Body mass index	Macau/Serbia	English	Constructed Response Expert	Uncertainty and data	Societal
CM038Q06	Body mass index	Macau/Serbia	English	Constructed Response Expert	Uncertainty and data	Societal

[Part 2/2]

Table A.5 **PISA 2012 Main Survey computer-based mathematics item classification**

Unit Item Code	Process	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
					Delta	Tau(1)	Tau(2)	1	2
CM002Q01	Employ	CM1	7.26	(0.15)	2.93353	-1.43002	1.43002	715.1	946.2
CM002Q02	Employ	CM1	32.82	(0.41)	0.11696			611.2	
CM004Q01T	Employ	CM1	40.60	(0.40)	-0.34452			575.2	
CM004Q02T	Employ	CM1	40.87	(0.40)	-0.43833			568.0	
CM004Q03	Formulate	CM1	44.42	(0.43)	-0.62056			553.7	
CM005Q01	Employ	CM4	32.40	(0.39)	0.05162			606.2	
CM005Q02	Employ	CM4	44.22	(0.41)	-0.57781			557.1	
CM005Q03	Employ	CM4	54.47	(0.43)	-1.10376			516.1	
CM005Q04	Employ	CM1	25.28	(0.29)	0.41056	-0.46088	0.46088	577.5	690.7
CM006Q01	Employ	CM1	66.49	(0.41)	-1.62347			475.6	
CM006Q02	Interpret	CM1	44.33	(0.26)	-0.47329	-1.49077	1.49077	445.5	685.0
CM006Q03	Formulate	CM1	22.41	(0.34)	0.57005			646.5	
CM008Q01	Formulate	CM1	23.51	(0.30)	0.31113	0.34202	-0.34202	599.2	653.4
CM008Q02	Formulate	CM3	9.49	(0.25)	1.42612			713.2	
CM011Q01T	Formulate	CM3	66.12	(0.41)	-1.66320			472.6	
CM011Q02	Employ	CM4	6.51	(0.20)	2.10146			765.8	
CM011Q03	Employ	CM4	11.92	(0.27)	1.49726			718.7	
CM014Q01	Employ	CM4	38.08	(0.40)	-0.22421			584.6	
CM014Q02	Employ	CM4	14.42	(0.29)	1.11078			688.6	
CM014Q03T	Employ	CM4	25.44	(0.38)	0.63770			651.8	
CM015Q01	Employ	CM4	59.02	(0.41)	-1.32981			498.5	
CM015Q02D	Formulate	CM2	8.43	(0.22)	1.17048	1.65678	-1.65678	685.8	700.7
CM015Q03D	Interpret	CM2	29.02	(0.33)	0.20264	-0.08940	0.08940	577.2	658.6
CM016Q01	Employ	CM2	64.71	(0.40)	-1.49186			485.9	
CM020Q01	Employ	CM2	29.58	(0.29)	0.31315	-0.61434	0.61434	562.1	690.9
CM020Q02	Employ	CM2	47.42	(0.44)	-0.67743			549.3	
CM020Q03	Employ	CM2	26.91	(0.35)	0.54081			644.2	
CM020Q04	Employ	CM2	44.12	(0.42)	-0.46689			565.7	
CM025Q01	Formulate	CM3	46.27	(0.42)	-0.56288			558.3	
CM025Q02T	Employ	CM3	58.36	(0.42)	-1.36826			495.5	
CM028Q03	Formulate	CM4	29.49	(0.39)	-0.04514			598.6	
CM035Q01	Interpret	CM3	78.41	(0.35)	-2.34025			419.8	
CM035Q02	Interpret	CM3	44.35	(0.41)	-0.46427			566.0	
CM035Q03	Interpret	CM3	65.72	(0.39)	-1.60875			476.9	
CM035Q04	Formulate	CM3	10.51	(0.25)	1.47715			717.2	
CM036Q01T	Interpret	CM3	43.81	(0.42)	-0.55090			559.2	
CM036Q02	Employ	CM3	9.08	(0.22)	1.92230			751.8	
CM036Q03	Interpret	CM3	11.11	(0.27)	1.64466			730.2	
CM038Q03T	Interpret	CM2	67.13	(0.40)	-1.71160			468.8	
CM038Q05	Interpret	CM2	27.75	(0.37)	0.50001			641.1	
CM038Q06	Interpret	CM2	23.24	(0.37)	0.74882			660.4	

[Part 1/2]

Table A.6 **PISA 2012 Main Survey problem solving item classification**

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Context (1)	Context (2)	Nature of Problem Situation
CP002Q06	Robot Cleaner	ACER	English	Constructed Response Expert	Technology	Social	Static
CP002Q07	Robot Cleaner	ACER	English	Simple Multiple Choice	Technology	Social	Static
CP002Q08	Robot Cleaner	ACER	English	Simple Multiple Choice	Technology	Social	Static
CP007Q01	Traffic	ACER	English	Simple Multiple Choice	Non-technology	Social	Static
CP007Q02	Traffic	ACER	English	Constructed Response Auto-coded	Non-technology	Social	Static
CP007Q03T	Traffic	ACER	English	Selected Response Variations	Non-technology	Social	Static
CP010Q01	New Furniture	ACER	English	Simple Multiple Choice	Non-technology	Personal	Static
CP010Q05	New Furniture	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Static
CP010Q06	New Furniture	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Static
CP014Q01	Lost	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Interactive
CP014Q02	Lost	ACER	English	Simple Multiple Choice	Non-technology	Personal	Interactive
CP014Q06	Lost	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Interactive
CP015Q01	Blocks Game	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Static
CP015Q02	Blocks Game	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Static
CP015Q04	Blocks Game	ACER	English	Constructed Response Auto-coded	Non-technology	Personal	Static
CP018Q04T	Communications	ETS	English	Constructed Response Expert	Non-technology	Social	Static
CP018Q05	Communications	ETS	English	Complex Multiple Choice	Non-technology	Social	Static
CP025Q01	Climate Control	University of Heidelberg	German	Constructed Response Auto-coded	Technology	Personal	Interactive
CP025Q02	Climate Control	University of Heidelberg	German	Constructed Response Auto-coded	Technology	Personal	Interactive
CP027Q01T	Aquarium	University of Heidelberg	German	Complex Multiple Choice	Non-technology	Personal	Interactive
CP027Q02	Aquarium	University of Heidelberg	German	Constructed Response Auto-coded	Non-technology	Personal	Interactive
CP028Q01	Advertising	ETS	English	Constructed Response Auto-coded	Non-technology	Social	Interactive
CP028Q02	Advertising	ETS	English	Constructed Response Auto-coded	Non-technology	Social	Interactive
CP029Q01	Remote Controls	University of Heidelberg	German	Constructed Response Auto-coded	Technology	Personal	Interactive
CP029Q02	Remote Controls	University of Heidelberg	German	Constructed Response Auto-coded	Technology	Personal	Interactive
CP032Q01	Drink Machine	ETS	English	Simple Multiple Choice	Technology	Social	Interactive
CP032Q02T	Drink Machine	ETS	English	Complex Multiple Choice	Technology	Social	Interactive
CP032Q04	Drink Machine	ETS	English	Simple Multiple Choice	Technology	Social	Interactive
CP034Q01T	Clock	ACER	English	Complex Multiple Choice	Technology	Personal	Interactive
CP034Q02	Clock	ACER	English	Constructed Response Auto-coded	Technology	Personal	Interactive
CP034Q05	Clock	ACER	English	Constructed Response Expert	Technology	Personal	Interactive
CP036Q01	Lights	ACER	English	Constructed Response Auto-coded	Technology	Personal	Interactive
CP036Q02	Lights	ACER	English	Constructed Response Expert	Technology	Personal	Static
CP036Q03	Lights	ACER	English	Constructed Response Expert	Technology	Personal	Interactive
CP037Q01	Vitamins	Sweden	English	Constructed Response Auto-coded	Non-technology	Social	Interactive
CP037Q02	Vitamins	Sweden	English	Selected Response Variations	Non-technology	Social	Interactive
CP037Q03	Vitamins	Sweden	English	Simple Multiple Choice	Non-technology	Social	Interactive
CP038Q01	Tickets	ACER	English	Constructed Response Auto-coded	Technology	Social	Interactive
CP038Q02	Tickets	ACER	English	Constructed Response Auto-coded	Technology	Social	Interactive
CP038Q03	Tickets	ACER	English	Constructed Response Auto-coded	Technology	Social	Interactive
CP041Q02	Smiley	University of Heidelberg	German	Constructed Response Expert	Technology	Personal	Interactive
CP041Q03	Smiley	University of Heidelberg	German	Constructed Response Auto-coded	Technology	Personal	Interactive

[Part 2/2]

Table A.6 **PISA 2012 Main Survey problem solving item classification**

Unit Item Code	Problem Solving Process	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
					Delta	Tau(1)	Tau(2)	1	2
CP002Q06	Representing and formulating	CP4	47.23	(0.26)	0.24755	-1.73723	1.73723	413.7	700.6
CP002Q07	Exploring and understanding	CP4	46.84	(0.43)	0.27462			559.3	
CP002Q08	Exploring and understanding	CP4	63.12	(0.42)	-0.58064			489.9	
CP007Q01	Planning and executing	CP3	86.44	(0.31)	-2.42254			340.4	
CP007Q02	Planning and executing	CP3	70.42	(0.40)	-1.11665			446.4	
CP007Q03T	Monitoring and reflecting	CP3	78.18	(0.37)	-1.59499			407.6	
CP010Q01	Exploring and understanding	CP3	75.59	(0.39)	-1.54065			412.0	
CP010Q05	Planning and executing	CP3	57.22	(0.45)	-0.37027			507.0	
CP010Q06	Planning and executing	CP3	42.49	(0.46)	0.34973			565.5	
CP014Q01	Planning and executing	CP4	56.07	(0.22)	-0.36172	-2.36313	2.36313	315.2	700.2
CP014Q02	Representing and formulating	CP4	22.24	(0.36)	1.67987			673.4	
CP014Q06	Representing and formulating	CP4	51.15	(0.45)	0.09181			544.5	
CP015Q01	Exploring and understanding	CP1	55.65	(0.45)	-0.18098			522.4	
CP015Q02	Planning and executing	CP1	39.96	(0.42)	0.59585			585.4	
CP015Q04	Monitoring and reflecting	CP1	9.99	(0.27)	2.74865			760.2	
CP018Q04T	Representing and formulating	CP2	16.82	(0.33)	1.98438			698.1	
CP018Q05	Exploring and understanding	CP2	43.25	(0.44)	0.42424			571.5	
CP025Q01	Representing and formulating	CP2	59.24	(0.41)	-0.36689	0.95308	-0.95308	491.7	522.8
CP025Q02	Planning and executing	CP2	27.92	(0.34)	1.16564	-0.02807	0.02807	591.6	671.8
CP027Q01T	Exploring and understanding	CP4	55.47	(0.35)	-0.17885	-0.48111	0.48111	462.5	582.6
CP027Q02	Planning and executing	CP4	16.18	(0.23)	2.35831	-0.98852	0.98852	638.7	818.1
CP028Q01	Representing and formulating	CP1	53.85	(0.43)	-0.17991			522.4	
CP028Q02	Planning and executing	CP1	61.32	(0.37)	-0.38544	-0.07416	0.07416	463.9	547.6
CP029Q01	Representing and formulating	CP3	65.68	(0.41)	-0.69594	0.85609	-0.85609	463.4	497.7
CP029Q02	Planning and executing	CP3	49.07	(0.41)	0.05188	0.15897	-0.15897	507.6	574.9
CP032Q01	Exploring and understanding	CP1	46.30	(0.45)	0.26577			558.6	
CP032Q02T	Representing and formulating	CP1	77.94	(0.30)	-1.54532	-0.60393	0.60393	345.1	478.1
CP032Q04	Planning and executing	CP1	50.01	(0.44)	-0.01101			536.2	
CP034Q01T	Exploring and understanding	CP1	40.30	(0.44)	0.62707			587.9	
CP034Q02	Planning and executing	CP1	92.26	(0.23)	-2.99980			293.6	
CP034Q05	Monitoring and reflecting	CP1	78.72	(0.37)	-1.54340			411.8	
CP036Q01	Representing and formulating	CP2	48.23	(0.45)	0.09071			544.4	
CP036Q02	Planning and executing	CP2	29.56	(0.28)	1.53249	-1.44876	1.44876	539.7	783.1
CP036Q03	Monitoring and reflecting	CP2	43.35	(0.45)	0.38056			568.0	
CP037Q01	Planning and executing	CP3	61.92	(0.45)	-0.69624			480.6	
CP037Q02	Exploring and understanding	CP3	44.07	(0.45)	0.22055			555.0	
CP037Q03	Monitoring and reflecting	CP3	16.18	(0.33)	1.86473			688.4	
CP038Q01	Exploring and understanding	CP2	50.11	(0.32)	0.10837	-1.03247	1.03247	453.3	638.4
CP038Q02	Planning and executing	CP2	58.02	(0.42)	-0.14067			525.6	
CP038Q03	Monitoring and reflecting	CP2	42.88	(0.44)	0.52118			579.3	
CP041Q02	Monitoring and reflecting	CP4	19.95	(0.33)	1.41837	0.76407	-0.76407	633.5	670.9
CP041Q03	Planning and executing	CP4	74.38	(0.23)	-2.09042	-2.35534	2.35534	175.5	559.2

[Part 1/2]

Table A.7 **PISA 2012 Main Survey financial literacy item classification**

Unit Item Code	Unit Name	Source	Language of submission	Item Format	Content	Process
PF001Q01	Costs of Running a Car	ACER	English	Complex Multiple Choice	Planning	Analyse information in a financial context
PF004Q03	Income tax	ACER	English	Constructed Response Expert	Planning	Evaluate financial issues
PF006Q02	Music system	ACER	English	Complex Multiple Choice	Planning	Analyse information in a financial context
PF009Q02	Shopping	ACER	English	Simple Multiple Choice	Money	Apply financial knowledge and understanding
PF010Q01	Bank statement	ACER	English	Constructed Response Manual	Money	Identify financial information
PF010Q02	Bank statement	ACER	English	Constructed Response Manual	Money	Analyse information in a financial context
PF012Q01	Interest	ACER	English	Simple Multiple Choice	RiskRew	Apply financial knowledge and understanding
PF012Q02	Interest	ACER	English	Complex Multiple Choice	RiskRew	Analyse information in a financial context
PF024Q02	Jacket sale	ACER	English	Constructed Response Expert	Money	Evaluate financial issues
PF028Q02	Phone plans	ACER	English	Constructed Response Expert	Planning	Analyse information in a financial context
PF028Q03	Phone plans	ACER	English	Simple Multiple Choice	Planning	Analyse information in a financial context
PF031Q01	Laptop	ACER	English	Complex Multiple Choice	RiskRew	Evaluate financial issues
PF031Q02	Laptop	ACER	English	Constructed Response Manual	RiskRew	Apply financial knowledge and understanding
PF033Q01	Wayne's Bank Statement	ACER	English	Simple Multiple Choice	Money	Analyse information in a financial context
PF033Q02	Wayne's Bank Statement	ACER	English	Simple Multiple Choice	Money	Identify financial information
PF035Q01	Ringtones	ACER	English	Constructed Response Manual	Landscape	Apply financial knowledge and understanding
PF036Q01	Online Shopping	ACER	English	Constructed Response Expert	Landscape	Evaluate financial issues
PF051Q01	Bicycle Shop	ACER	English	Constructed Response Expert	Planning	Evaluate financial issues
PF051Q02	Bicycle Shop	ACER	English	Constructed Response Expert	Planning	Evaluate financial issues
PF052Q01	Video Game	ACER	English	Complex Multiple Choice	Planning	Identify financial information
PF054Q01	E-mail	ACER	English	Constructed Response Expert	Landscape	Evaluate financial issues
PF055Q03	Invoice	ACER	English	Constructed Response Manual	Money	Apply financial knowledge and understanding
PF058Q01	Personal Identification Number	ACER	English	Constructed Response Expert	RiskRew	Evaluate financial issues
PF062Q01	Mobile Phone Contract	ACER	English	Complex Multiple Choice	Landscape	Evaluate financial issues
PF068Q01	Job Change	ACER	English	Constructed Response Expert	Planning	Evaluate financial issues
PF069Q01	Student Account	ACER	English	Complex Multiple Choice	Landscape	Analyse information in a financial context
PF075Q02	Study Options	ACER	English	Complex Multiple Choice	Planning	Analyse information in a financial context
PF082Q01	New Bike	ACER	English	Constructed Response Expert	Money	Identify financial information
PF082Q02	New Bike	ACER	English	Simple Multiple Choice	RiskRew	Identify financial information
PF095Q01	Changing Value	ACER	English	Simple Multiple Choice	Money	Identify financial information
PF095Q02	Changing Value	ACER	English	Complex Multiple Choice	Landscape	Analyse information in a financial context
PF097Q01	Company Profit	ACER	English	Complex Multiple Choice	Landscape	Identify financial information
PF102Q01	Gantica	ACER	English	Constructed Response Manual	RiskRew	Apply financial knowledge and understanding
PF102Q02	Gantica	ACER	English	Constructed Response Expert	RiskRew	Apply financial knowledge and understanding
PF103Q01	Investing	ACER	English	Constructed Response Expert	RiskRew	Evaluate financial issues
PF105Q01	Interest Rates	ACER	English	Simple Multiple Choice	Money	Apply financial knowledge and understanding
PF105Q02	Interest Rates	ACER	English	Simple Multiple Choice	Money	Apply financial knowledge and understanding
PF106Q01	Family Holiday	ACER	English	Constructed Response Expert	Planning	Evaluate financial issues
PF106Q02	Family Holiday	ACER	English	Simple Multiple Choice	Planning	Apply financial knowledge and understanding
PF110Q01	Living Alone	ACER	English	Complex Multiple Choice	Planning	Evaluate financial issues

[Part 2/2]

Table A.7 **PISA 2012 Main Survey financial literacy item classification**

Unit Item Code	Context	Cluster	International % correct	S.E. % correct	Item parameters (RP=0.50)			Thresholds (RP=0.62) PISA scale	
					Delta	Tau(1)	Tau(2)	1	2
PF001Q01	Home and family	PF1	74.81	(0.37)	-1.02000			426.8	476.2
PF004Q03	Education and work	PF2	6.98	(0.21)	3.29700			829.1	
PF006Q02	Individual	PF2	57.67	(0.43)	-0.08100			514.3	
PF009Q02	Home and family	PF1,PFUH	93.17	(0.23)	-2.76900			263.8	
PF010Q01	Home and family	PF1	50.40	(0.45)	0.34300			553.8	
PF010Q02	Home and family	PF1	25.59	(0.39)	1.29300	0.78400	-0.78400	621.2	663.4
PF012Q01	Individual	PF1	60.74	(0.43)	-0.18400			504.7	
PF012Q02	Individual	PF1	47.56	(0.43)	0.51300			569.6	
PF024Q02	Individual	PF2	62.54	(0.41)	-0.35000			489.1	
PF028Q02	Individual	PF1,PFUH	59.00	(0.43)	-0.19800			503.4	
PF028Q03	Individual	PF1,PFUH	74.90	(0.38)	-1.03700			425.2	
PF031Q01	Home and family	PF1,PFUH	29.08	(0.40)	1.36600			649.1	
PF031Q02	Home and family	PF1,PFUH	56.47	(0.43)	-0.01000			520.8	
PF033Q01	Individual	PF2,PFUH	37.75	(0.41)	0.97900			613.1	
PF033Q02	Individual	PF2,PFUH	58.62	(0.43)	-0.05600			516.6	
PF035Q01	Individual	PF2,PFUH	53.59	(0.43)	0.19200			539.7	
PF036Q01	Societal	PF1	46.80	(0.42)	0.52200			570.5	
PF051Q01	Education and work	PF2	83.66	(0.34)	-1.54100			378.2	
PF051Q02	Education and work	PF2	48.85	(0.45)	0.33400			552.9	
PF052Q01	Individual	PF2	76.92	(0.38)	-1.09800			419.5	
PF054Q01	Societal	PF1	66.79	(0.40)	-0.53600			471.9	
PF055Q03	Individual	PF2	37.58	(0.37)	0.82100	-0.25800	0.25800	541.6	655.1
PF058Q01	Societal	PF2,PFUH	86.55	(0.29)	-1.66800			366.5	
PF062Q01	Home and family	PF2,PFUH	75.64	(0.39)	-0.82700			444.8	
PF068Q01	Education and work	PF1	52.51	(0.44)	0.22200			542.5	
PF069Q01	Education and work	PF2	69.53	(0.41)	-0.67600			458.7	
PF075Q02	Education and work	PF2	31.07	(0.39)	1.23300			636.7	
PF082Q01	Individual	PF1	65.93	(0.40)	-0.36000	0.75100	-0.75100	466.5	510.1
PF082Q02	Home and family	PF1	83.96	(0.31)	-1.69200			364.1	
PF095Q01	Home and family	PF2	33.17	(0.43)	1.05400			620.0	
PF095Q02	Societal	PF2	28.03	(0.39)	1.44800			656.8	
PF097Q01	Individual	PF1	11.27	(0.28)	2.76000			779.1	
PF102Q01	Home and family	PF2,PFUH	85.51	(0.32)	-1.72700			360.9	
PF102Q02	Home and family	PF2,PFUH	65.43	(0.33)	-0.44800	-0.39100	0.39100	416.3	543.9
PF103Q01	Individual	PF1	35.25	(0.44)	1.05200			619.9	
PF105Q01	Individual	PF1,PFUH	33.46	(0.43)	1.17900			631.7	
PF105Q02	Individual	PF1,PFUH	42.99	(0.45)	0.73400			590.2	
PF106Q01	Home and family	PF2,PFUH	77.05	(0.38)	-1.11200			418.2	
PF106Q02	Home and family	PF2,PFUH	54.87	(0.44)	0.17300			537.9	
PF110Q01	Home and family	PF1,PFUH	90.91	(0.26)	-2.12600			323.7	

ANNEX B – CONTRAST CODING USED IN CONDITIONING

[Part 1/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Grade	ST01Q01	7-14 Ungraded Missing	value – mode 0 0 0 0 0 0 1 1
Study programme	ST02Q01	National categories	If there is at least one school with more than one study programme in a country, national study programmes are dummy coded with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Age of student	AGE	Value (decimal) Missing	value – median 0 0 1
Gender	ST04Q01	1 Female 2 Male Missing	Two dummies if missing data is present and one dummy if no missing data with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
ISCED 0	ST05Q01	1 No 2 Yes, one year or less 3 yes, more than one year Missing (or invalid)	Three dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Age when started ISCED 1	ST06Q01	Value Missing	value – median 0 0 1
Repeated grade at ISCED 1	ST07Q01	1 No 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Three dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Repeated grade at ISCED 2	ST07Q02	1 No 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Three dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Repeated grade at ISCED 3	ST07Q03	1 No 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Three dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Arrived late for school	ST08Q01	1 None 2 One or two times 3 Three or four times 4 Five or more times Missing (or invalid)	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Skipped the whole school day	ST09Q01	1 None 2 One or two times 3 Three or four times 4 Five or more times Missing (or invalid)	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Skipped some classes	ST115Q01	1 None 2 One or two times 3 Three or four times 4 Five or more times Missing (or invalid)	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Mother	ST11Q01	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Father	ST11Q02	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Brother(s)	ST11Q03	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Sister(s)	ST11Q04	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Grandparent(s)	ST11Q05	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Lives at home with you – Other(s)	ST11Q06	1 Yes 2 No Missing (or invalid)	Two dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Occupational status of Mother (SEI)	BMMJ1	16-90 (decimal) Missing	value – median 0 0 1
Occupational status of Father (SEI)	BFMJ2	16-90 (decimal) Missing	value-median 0 0 1
Educational level of Mother (MISCED)	ST13Q01 ST14Q01 ST14Q02 ST14Q03 ST14Q04	5 None 4 ISCED 1 3 ISCED 2 2 ISCED 3B, C 1 ISCED 3A, Missing 1 Yes 2 No Missing	Item ST13Q01 was recoded as (5=0), (4=1), (3=2), (2=3), (3=4). Item ST14Q04 was recoded as (1=4), (2=0) Item ST14Q03 was recoded as (1=5), (2=0) Item ST14Q02 was recoded as (1=5), (2=0) Item ST14Q01 was recoded as (1=6), (2=0). New variable MISCED was created as maximum value of five items, thus having categories from 0 to 6. Plus one category for missing (when all five items are missing) Seven dummy variables were created based on the value of MISCED and with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)

[Part 2/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Educational level of Father (FISCED)	ST17Q01 ST18Q01 ST18Q02 ST18Q03 ST18Q04	5 None 4 ISCED 1 3 ISCED 2 2 ISCED 3B, C 1 ISCED 3A, Missing 1 Yes 2 No Missing	Item ST17Q01 was recoded as (5=0), (4=1), (3=2), (2=3), (3=4). Item ST18Q04 was recoded as (1=4), (2=0) Item ST18Q03 was recoded as (1=5), (2=0) Item ST18Q02 was recoded as (1=5), (2=0) Item ST18Q01 was recoded as (1=6), (2=0). New variable FISCED was created as maximum value of five items, thus having categories from 0 to 6. Plus one category for missing (when all five items are missing) Seven dummy variables were created based on the value of FISCED and with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
What Mother is currently doing	ST15Q01	1 Working full-time 2 Working part-time 3 Not working, looking 4 Other Missing (or invalid)	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
What Father is currently doing	ST19Q01	1 Working full-time 2 Working part-time 3 Not working, looking 4 Other Missing (or invalid)	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Immigration status (IMMIG)	ST20int (CTSELF) (CTFATHER) (CTMOTHER)	1 Native 2 Second-Generation 3 First-Generation Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Country arrival age	ST21Q01	Value N/A (born in country) Missing (or >17)	(copy) 0 0 0 0 -1
Language at home	ST25int	1 Language of test 2 Other language Missing	-1 -1 01 00 00 01
Family wealth (WEALTH)	ST26Q02 ST26Q06 ST26Q13 ST26Q14 ST26Q15 ST26Q16 ST26Q17 ST27Q01 ST27Q02 ST27Q03 ST27Q04 ST27Q05	1 Yes 2 No Missing 1 None 2 One 3Two 4Three or more Missing	All items of Q26 were recoded as (Yes=1, No=0) and all items of Q27 were recoded as (1=0, 2=1, 3=2, 4=3). Total score was calculated as a ratio of a sum of all items over maximum score of valid responses (items with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Home educational resources (HEDRES)	ST26Q01 ST26Q03 ST26Q04 ST26Q05 ST26Q10 ST26Q11 ST26Q12	1 Yes 2 No Missing	All items were recoded as (Yes=1, No=0). Total score was calculated as a ratio of a sum of all items over maximum score of valid responses (items with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Cultural possessions at home (CULTPOS)	ST26Q07 ST26Q08 ST26Q09	1 Yes 2 No Missing	All items were recoded as (Yes=1, No=0). Total score was calculated as a ratio of a sum of all items over maximum score of valid responses (items with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
How many books at home	ST28Q01	1 0-10 books 2 11-25 books 3 26-100 books 4 101-200 books 5 201-500 books 6 More than 500 books Missing	Six dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Mathematics interest (INTMAT)	ST29Q01 ST29Q03 ST29Q04 ST29Q06	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (4=0),(3=1),(2=2),(1=3).Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Instrumental motivation for mathematics (INSTMOT)	ST29Q02 ST29Q05 ST29Q07 ST29Q08	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (4=0), (3=1), (2=2), (1=3).Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Subjective norms in mathematics (SUBNORM)	ST35Q01 ST35Q02 ST35Q03 ST35Q04 ST35Q05 ST35Q06	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (4=0), (3=1), (2=2), (1=3).Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics self-efficacy (MATHEFF)	ST37Q01 ST37Q02 ST37Q03 ST37Q04 ST37Q05 ST37Q06 ST37Q07 ST37Q08	1 Very confident 2 Confident 3 Not very confident 4 Not at all confident Missing	Items were reversely recoded as (4=0), (3=1), (2=2), (1=3).Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 3/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Mathematics anxiety (ANXMAT)	ST42Q01 ST42Q03 ST42Q05 ST42Q08 ST42Q10	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (4=0), (3=1), (2=2), (1=3).Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics self-concept (SCMAT)	ST42Q02 ST42Q04 ST42Q06 ST42Q07 ST42Q09	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items except item 02 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Item 02 was coded as (1=0), (2=1), (3=2), (4=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Perceived control of success in mathematics	ST43Q01 ST43Q02 ST43Q03 ST43Q04 ST43Q05 ST43Q06	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items 01, 02, 05 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Items 03, 04, 06 were coded as (1=0), (2=1), (3=2), (4=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Attributions to failure in mathematics (FAILMAT)	ST44Q01 ST44Q03 ST44Q04 ST44Q05 ST44Q07 ST44Q08	1 Very Likely 2 Likely 3 Slightly likely 4 Not at all likely Missing	Items were coded as (1=0), (2=1), (3=2), (4=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics work ethic (MATWKETH)	ST46Q01 ST46Q02 ST46Q03 ST46Q04 ST46Q05 ST46Q06 ST46Q07 ST46Q08 ST46Q09	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics intentions (MATINTFC)	ST48Q01	1 Courses after school -Maths 2 Courses after school -Test Language Missing	All items were reversely recoded as (2=0), (1=1). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
	ST48Q02	1 Major in college - Math 2 Major in college - Science Missing	
	ST48Q03	1 Study harder - Math 2 Study harder - Test Language Missing	
	ST48Q04	1 Maximum classes - Math 2 Maximum classes - Science Missing	
	ST48Q05	1 Pursuing a career - Math 2 Pursuing a career -Science Missing	
Mathematics behaviour (MATBEH)	ST49Q01 ST49Q02 ST49Q03 ST49Q04 ST49Q05 ST49Q06 ST49Q07 ST49Q09	1 Always or almost always 2 Often 3 Sometimes 4 Never or rarely Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Learning strategies	ST53Q01 ST53Q02 ST53Q03 ST53Q04	Choice of three strategies 1 First strategy 2 second strategy 3 third strategy No response (Missing)	Items 01 and 02 were recoded as ('1'=1), ('2'=0), ('3'=0). Item 03 was recoded as ('1'=0), ('2'=1), ('3'=0). Item 04 was recoded as ('1'=0), ('2'=0), ('3'=1). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Out of school lessons in - test language (hours a week)	ST55Q01	1 'I do not attend' 2 'Less than 2 hours' 3 'Between 2 and 4 hours' 4 'Between 4 and 6 hours' 5 '6 or more hours' Missing	The item was recoded as ('1'=0), ('2'=1), ('3'=3), ('4'=5), ('5'=7), representing an approximate time in hours. Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Out of school lessons in mathematics (hours)	ST55Q02	Value Missing	The item was recoded as ('1'=0), ('2'=1), ('3'=3), ('4'=5), ('5'=7), representing an approximate time in hours. Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Out of school lessons in - science (hours)	ST55Q03	Value Missing	The item was recoded as ('1'=0), ('2'=1), ('3'=3), ('4'=5), ('5'=7), representing an approximate time in hours. Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Out of school lessons in - other subjects (hours)	ST55Q04	Value Missing	The item was recoded as ('1'=0), ('2'=1), ('3'=3), ('4'=5), ('5'=7), representing an approximate time in hours. Two dummy variable were created as follows: Value value – median 0 Missing 0 1

[Part 4/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Homework set by teacher (hours)	ST57Q01	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
How many hour of homework with somebody overlooking and helping	ST57Q02	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Work with a personal <tutor>	ST57Q03	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Classes through a commercial company	ST57Q04	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Study with parents or family	ST57Q05	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Train school lessons on a computer	ST57Q06	Value Missing	Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Experience with applied mathematics tasks at school (EXAPPLM)	ST61Q01 ST61Q02 ST61Q03 ST61Q04 ST61Q06 ST61Q08	1 Frequently 2 Sometimes 3 Rarely 4 Never Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Experience with pure mathematics tasks at school (EXPUREM)	ST61Q05 ST61Q07 ST61Q09	1 Frequently 2 Sometimes 3 Rarely 4 Never Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Experience with applied mathematics tasks at school (FAMCON)	ST62Q01 ST62Q02 ST62Q03 ST62Q06 ST62Q07 ST62Q08 ST62Q09 ST62Q10 ST62Q12 ST62Q15 ST62Q16 ST62Q17 ST62Q19	1 Never heard of it 2 Heard of it once or twice 3 Heard of it a few times 4 Heard of it often 5 Know it well Missing	All items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Experience with applied mathematics tasks at school (FAMCONC)	ST62Q04 ST62Q11 ST62Q12	1 Never heard of it 2 Heard of it once or twice 3 Heard of it a few times 4 Heard of it often 5 Know it well Missing	All items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Average time per week on <language> (LMINS)	ST69Q01 ST70Q01	Value Missing	The value is the product of ST69Q01*ST70Q01. Two dummy variable were created as follows: Value value – mean 0 Missing 0 1
Average time per week on mathematics (MMINS)	ST69Q02 ST70Q02	Value Missing	The value is the product of ST69Q02*ST70Q02. Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Average time per week on science (SMINS)	ST69Q03 ST70Q03	Value Missing	The value is the product of ST69Q03*ST70Q03. Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Total number of classes per week	ST71Q01	Value Missing	value – median 0 0 1
Number of students attending <language> class	ST72Q01	Value Missing	value – median 0 0 1
Experience with these types of problems at school	ST73Q01 ST73Q02 ST74Q01 ST74Q02 ST75Q01 ST75Q02 ST76Q01 ST76Q02	1 Frequently 2 Sometimes 3 Rarely 4 Never Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Teacher support (TEACHSUP)	ST77Q01 ST77Q02 ST77Q04 ST77Q05 ST77Q06	1 Every lesson 2 Most lessons 3 Some lessons 4 Never or hardly ever Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Teacher behaviour: teacher-directed instruction (TCHBEHTD)	ST79Q01 ST79Q02 ST79Q06 ST79Q08 ST79Q15	1 Every lesson 2 Most lessons 3 Some lessons 4 Never or hardly ever Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Teacher behaviour: student orientation (TCHBEHSO)	ST79Q03 ST79Q04 ST79Q07 ST79Q10	1 Every lesson 2 Most lessons 3 Some lessons 4 Never or hardly ever Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 5/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Teacher behaviour: formative assessment (TCHBEHFA)	ST79Q05 ST79Q11 ST79Q12 ST79Q17	1 Every lesson 2 Most lessons 3 Some lessons 4 Never or hardly ever Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Cognitive activation in mathematics lessons (COGACT)	ST80Q01 ST80Q04 ST80Q05 ST80Q06 ST80Q07 ST80Q08 ST80Q09 ST80Q10 ST80Q11	1 Always or almost always 2 Often 3 Sometimes 4 Never or rarely Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Disciplinary climate (DISCLIMA)	ST81Q01 ST81Q02 ST81Q03 ST81Q04 ST81Q05	1 Every lesson 2 Most lessons 3 Some lessons 4 Never or hardly ever Missing	All items were coded as (1=0), (2=1), (3=2), (4=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics teacher's support anchoring vignettes (ANCMTSUP)	ST82Q01 ST82Q02 ST82Q03	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics teacher's support (MTSUP)	ST83Q01 ST83Q02 ST83Q03	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics teacher's classroom management anchoring vignettes (ANCCLSMAN)	ST84Q01 ST84Q02 ST84Q03	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics teacher's classroom management (CLSMAN)	ST85Q01 ST85Q02 ST85Q03 ST85Q04	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items 01, 02, 03 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Item 04 was coded as (4=3), (3=2), (2=1), (1=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Teacher-student relations (STUDREL)	ST86Q01 ST86Q02 ST86Q03 ST86Q04 ST86Q05	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Sense of belonging to school (BELONG)	ST87Q01 ST87Q02 ST87Q03 ST87Q04 ST87Q05 ST87Q06 ST87Q07 ST87Q08 ST87Q09	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items 01, 04 and 06 were coded as (4=3), (3=2), (2=1), (1=0). Items 02, 03, 05, 07, 08 and 09 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Attitude towards school: learning outcomes (ATSCHL)	ST88Q01 ST88Q02 ST88Q03 ST88Q04	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items 03 and 04 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Items 01 and 02 were coded as (4=3), (3=2), (2=1), (1=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Attitude towards school: learning activities (ATTLNACT)	ST89Q02 ST89Q03 ST89Q04 ST89Q05	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Perceived control of success at school	ST91Q01 ST91Q02 ST91Q03 ST91Q04 ST91Q05 ST91Q06	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items 01, 02 and 05 were reversely recoded as (4=0), (3=1), (2=2), (1=3). Items 03, 04 and 06 were coded as (4=3), (3=2), (2=1), (1=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Students' perseverance (PERSEV)	ST93Q01 ST93Q03 ST93Q04 ST93Q06 ST93Q07	1 Very much like me 2 Mostly like me 3 Somewhat like me 4 Not much like me 5 Not at all like me Missing	Items 04, 06 and 07 were reversely recoded as (5=0), (4=1), (3=2), (2=3), (1=4). Items 01 and 03 were coded as (5=4), (4=3), (3=2), (2=1), (1=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Openness for problem solving (OPENPS)	ST94Q05 ST94Q06 ST94Q09 ST94Q10 ST94Q14	1 Very much like me 2 Mostly like me 3 Somewhat like me 4 Not much like me 5 Not at all like me Missing	All items were reversely recoded as (5=0), (4=1), (3=2), (2=3), (1=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 6/6]

Table B.1 PISA 2012 Main Survey contrast coding used in conditioning for the student questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
STUDENT QUESTIONNAIRE			
Problem-sloving strategy: systematic strategies	ST96Q02 ST101Q01 ST101Q02 ST104Q01	1 I would definitely do this 2 I would probably do this 3 I would probably not do this 4 I would definitely not do this Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Problem-sloving strategy: unsystematic strategies	ST96Q01 ST101Q03 ST101Q05 ST104Q04	1 I would definitely do this 2 I would probably do this 3 I would probably not do this 4 I would definitely not do this Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Problem-sloving strategy: seeking help	ST96Q03 ST96Q05 ST104Q05 ST104Q06	1 I would definitely do this 2 I would probably do this 3 I would probably not do this 4 I would definitely not do this Missing	All items were reversely recoded as (4=0), (3=1), (2=2), (1=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 1/2]

Table B.2 PISA 2012 Main Survey contrast coding used in conditioning for the ICT questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
ICT QUESTIONNAIRE			
ICT availability at home (ICTHOME)	IC01Q01 IC01Q02 IC01Q03 IC01Q04 IC01Q05 IC01Q06 IC01Q07 IC01Q08 IC01Q09 IC01Q10 IC01Q11	1 Yes, and I use it 2 Yes, but I don't use it 3 No Missing	Items were reversely recoded as (3=0), (2=1), (1=2). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
ICT availability at school (ICTSCH)	IC02Q01 IC02Q02 IC02Q03 IC02Q04 IC02Q05 IC02Q06 IC02Q07	1 Yes, and I use it 2 Yes, but I don't use it 3 No Missing	Items were reversely recoded as (3=0), (2=1), (1=2). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
How old were you when you first used a computer	IC03Q01	1 6 y/o or younger 2 Between 7 and 9 y/o 3 Between 10 and 12 y/o 4 13 y/o or older 5 Never used Missing (or invalid)	Five dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
How old were you when you first accessed the internet	IC04Q01	1 6 y/o or younger 2 Between 7 and 9 y/o 3 Between 10 and 12 y/o 4 13 y/o or older 5 Never used Missing (or invalid)	Five dummies with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Internet use at school on a typical weekday	IC05Q01	01 No time 02 1-30 minutes 03 31-60 minutes 04 Between 1 and 2 hours 05 Between 2 and 4 hours 06 Between 4 and 6 hours 07 More than 6 hours Missing (or invalid)	Items were recoded to represent numerical value in half-hours as follows ('01'=0), ('02'=1), ('03'=2), ('04'=4), ('05'=8), ('06'=12), ('07'=18). Two dummy variables were created as follows: Value value – median 0 Missing 0 1
Internet use outside of school on a typical weekday	IC06Q01	01 No time 02 1-30 minutes 03 31-60 minutes 04 Between 1 and 2 hours 05 Between 2 and 4 hours 06 Between 4 and 6 hours 07 More than 6 hours Missing (or invalid)	Items were recoded to represent numerical value in half-hours as follows ('01'=0), ('02'=1), ('03'=2), ('04'=4), ('05'=8), ('06'=12), ('07'=18). Two dummy variables were created as follows: Value value – median 0 Missing 0 1
Internet use on a typical weekend day	IC07Q01	01 No time 02 1-30 minutes 03 31-60 minutes 04 Between 1 and 2 hours 05 Between 2 and 4 hours 06 Between 4 and 6 hours 07 More than 6 hours Missing (or invalid)	Items were recoded to represent numerical value in half-hours as follows ('01'=0), ('02'=1), ('03'=2), ('04'=4), ('05'=8), ('06'=12), ('07'=18). Two dummy variables were created as follows: Value value – median 0 Missing 0 1
ICT entertainment use (ENTUSE)	IC08Q01 IC08Q02 IC08Q03 IC08Q04 IC08Q05 IC08Q06 IC08Q07 IC08Q08 IC08Q09 IC08Q11	1 Never or hardly ever 2 Once or twice a month 3 Once or twice a week 4 Almost every day 5 Every day Missing	Items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 2/2]

Table B.2 PISA 2012 Main Survey contrast coding used in conditioning for the ICT questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
ICT QUESTIONNAIRE			
ICT use at home for school related tasks (HOMSCH)	IC09Q01 IC09Q02 IC09Q03 IC09Q04 IC09Q05 IC09Q06 IC09Q07	1 Never or hardly ever 2 Once or twice a month 3 Once or twice a week 4 Almost every day 5 Every day Missing	Items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Use of ICT for school (USESCH)	IC10Q01 IC10Q02 IC10Q03 IC10Q04 IC10Q05 IC10Q06 IC10Q07 IC10Q08 IC10Q09	1 Never or hardly ever 2 Once or twice a month 3 Once or twice a week 4 Almost every day 5 Every day Missing	Items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Use of computer in mathematics lessons by students (USEMATH1)	IC11Q01 IC11Q02 IC11Q03 IC11Q04 IC11Q05 IC11Q06 IC11Q07	1 Yes, students did this 2 Yes, but only the teacher demonstrated this 3 No Missing	Items were recoded as (1=1), (2=0), (3=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Use of computer in mathematics lessons by teacher demonstrating only (USEMATH2)	IC11Q01 IC11Q02 IC11Q03 IC11Q04 IC11Q05 IC11Q06 IC11Q07	1 Yes, students did this 2 Yes, but only the teacher demonstrated this 3 No Missing	Items were recoded as (1=0), (2=1), (3=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Attitudes towards computers: computer as a tool for school learning (ICTATTPOS)	IC22Q01 IC22Q02 IC22Q04	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Attitudes towards computers: limitations of the computer as a tool for school learning (ICTATTNEG)	IC22Q06 IC22Q07 IC22Q08	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were coded as (1=0), (2=1), (3=2), (4=3). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 1/3]

Table B.3 PISA 2012 Main Survey contrast coding used in conditioning for the educational career questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
EDUCATIONAL CAREER QUESTIONNAIRE			
Did you ever miss two or more consecutive months of ISCED 1	EC01Q01	1 No, never 2 Yes, once 3 Yes, twice or more Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Did you ever miss two or more consecutive months of ISCED 2	EC02Q01	1 No, never 2 Yes, once 3 Yes, twice or more Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Information about careers (INFOCAR)	EC03Q01 EC03Q02 EC03Q03 EC03Q04 EC03Q05 EC03Q06 EC03Q07 EC03Q08 EC03Q09 EC03Q10	1 Yes 2 No, never Missing	Items were reversely recoded as (1=1), (2=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Skills acquired at school (INFOJOB1) – How to find info on jobs – How to search for a job – How to write a resume – How to prepare for a job interview – How to find info on <ISCED 3-5> – How to find info on student grants	EC04Q01A EC04Q02A EC04Q03A EC04Q04A EC04Q05A EC04Q06A	1 Tick 2 No tick	Items were reversely recoded as (1=1), (2=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Skills acquired out of school (INFOJOB2) – How to find info on jobs – How to search for a job – How to write a resume – How to prepare for a job interview – How to find info on <ISCED 3-5> – How to find info on student grants	EC04Q01B EC04Q02B EC04Q03B EC04Q04B EC04Q05B EC04Q06B	1 Tick 2 No tick	Items were reversely recoded as (1=1), (2=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 2/3]

Table B.3 PISA 2012 Main Survey contrast coding used in conditioning for the educational career questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
EDUCATIONAL CAREER QUESTIONNAIRE			
The first language learned at home	EC05Q01	1 Test language or other official dialect 2 Test language or other official dialect and another language 3 Language other than the test or other official language Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Age when started learning test language	EC06Q01	1 Between 0 and 3 y/o 2 Between 4 and 6 y/o 3 Between 7 and 9 y/o 4 Between 10 and 12 y/o 5 13 y/o or older Missing (or invalid)	Items were recoded to represent numerical value in years as follows ('1'=1), ('2'=5), ('3'=8), ('4'=11), ('5'=14). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Language spoken with my mother	EC07Q01	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language spoken with my father	EC07Q02	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language spoken with my brother(s)/sister(s)	EC07Q03	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language spoken with my best friend	EC07Q04	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language spoken with my schoolmates	EC07Q05	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language usually used for reading books, magazines, newspapers	EC08Q01	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language usually used for watching TV or movies	EC08Q02	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language usually used for surfing the internet	EC08Q03	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Language usually used for writing e-mails or letters	EC08Q04	1 Mostly my heritage language 2 About equally heritage and test languages 3 Mostly test language 4 Not applicable Missing	Four dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Attended <remedial lessons> in test language	EC09Q03	1 Yes 2 No, never Missing	Two dummy variables with default value of '00' and - national mode = '-1' in both dummies - corresponding category = '01' (including missing)
Attended lessons in my heritage language	EC11Q02	1 Yes 2 No, never Missing	Two dummy variables with default value of '00' and - national mode = '-1' in both dummies - corresponding category = '01' (including missing)
Attended instruction in school subjects through my heritage language	EC11Q03	1 Yes 2 No, never Missing	Two dummy variables with default value of '00' and - national mode = '-1' in both dummies - corresponding category = '01' (including missing)
Hours per week receiving systematic support for improving test language	EC10Q01	1 None 2 Less than 2 3 2 or more but less than 4 4 4 or more but less than 6 5 6 or more Missing (or invalid)	Items were coded to represent numerical value in hours per week as follows ('1'=0), ('2'=1), ('3'=2), ('4'=3), ('5'=5). Two dummy variables were created as follows: Value value – median 0 Missing 0 1

[Part 3/3]

Table B.3 PISA 2012 Main Survey contrast coding used in conditioning for the educational career questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
EDUCATIONAL CAREER QUESTIONNAIRE			
Hours per attending either lessons in heritage language or instructions through heritage language	EC12Q01	1 None 2 Less than 2 3 2 or more but less than 4 4 4 or more but less than 6 5 6 or more Missing (or invalid)	Items were coded to represent numerical value in hours per week as follows ('1'=0), ('2'=1), ('3'=2), ('4'=3), ('5'=5). Two dummy variables were created as follows: Value value – median 0 Missing 0 1
Mother born in country of test	ST22Q01	1 No 2 Yes Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Acculturation: host culture oriented strategies (HOSTCUL)	ST23Q01 ST23Q03 ST23Q05 ST23Q07	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Acculturation: heritage culture oriented strategies (HERITCUL)	ST23Q02 ST23Q04 ST23Q06 ST23Q08	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Cultural distance between host and heritage culture (CULTDIST)	ST24Q01 ST24Q02 ST24Q03	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	Items were reversely recoded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1

[Part 1/3]

Table B.4 PISA 2012 Main Survey contrast coding used in conditioning for the parent questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
PARENT QUESTIONNAIRE			
Who will complete this questionnaire – Mother or female guardian	PA01Q01	1 Tick 2 No tick Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Who will complete this questionnaire – Father or male guardian	PA01Q02	1 Tick 2 No tick Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Who will complete this questionnaire – other	PA01Q03	1 Tick 2 No tick Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Age of Father	PA02Q01	1 Younger than 36 y/o 2 Between 36 and 40 y/o 3 Between 41 and 45 y/o 4 Between 46 and 50 y/o 5 51 y/o or older Missing (or invalid)	Items were coded as follows ('1'=0), ('2'=1), ('3'=2), ('4'=3), ('5'=5). Two dummy variables were created as follows: Value value – median 0 Missing 0 1
Does the child's Father have any of the following qualifications - ISCED 5A, 6 - ISCED 5B - ISCED 4 - ISCED 3A (PQFISCED)	PA03Q01 PA03Q02 PA03Q03 PA03Q04	1 Yes 2 No Missing	Item PA03Q04 was recoded as (1=1), (2=0) Item PA03Q03 was recoded as (1=1), (2=0) Item PA03Q02 was recoded as (1=2), (2=0) Item PA03Q01 was recoded as (1=3), (2=0). New variable PQFISCED was created as maximum value of four items, thus having categories from 0 to 3. Plus one category for missing (when all four items are missing) Four dummy variables were created based on the value of PQFISCED and with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Occupational status Father, parents answer (SEI)	BFMJ4 (based on PA04Q01)	16-90(decimal) Missing	value – median 0 0 1
Does the child's Mother have any of the following qualifications - ISCED 5A, 6 - ISCED 5B - ISCED 4 - ISCED 3A (PQMISCED)	PA05Q01 PA05Q02 PA05Q03 PA05Q04	1 Yes 2 No Missing	Item PA05Q04 was recoded as (1=1), (2=0) Item PA05Q03 was recoded as (1=1), (2=0) Item PA05Q02 was recoded as (1=2), (2=0) Item PA05Q01 was recoded as (1=3), (2=0). New variable PQMISCED was created as maximum value of four items, thus having categories from 0 to 3. Plus one category for missing (when all four items are missing) Four dummy variables were created based on the value of PQMISCED and with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Occupational status Mother, parents answer (SEI)	BMMJ3 (based on PA06Q01)	16-90(decimal) Missing	value-median 0 0 1
Annual household income	PA07Q01	1 Less than <$A> 2 <$A> or more and less <$B> 3 <$B> or more and less <$C> 4 <$C> or more and less <$D> 5 <$D> or more and less <$E> 6 <$E> or more Missing	Items were coded as (6=5), (5=4), (4=3), (3=2), (2=1), (1=0). Two dummy variable were created as follows: Value value – median 0 Missing 0 1

[Part 2/3]

Table B.4 **PISA 2012 Main Survey contrast coding used in conditioning for the parent questionnaire variables**

Variable	Variable name	Variable coding	Contrast coding
PARENT QUESTIONNAIRE			
Paid to educational providers in the last year	PA08Q01	1 Nothing 2 More than $0 and less <$W> 3 <$W> or more and less <$X> 4 <$X> or more and less <$Y> 5 <$Y> or more and less <$Z> 6 <$Z> or more Missing	Items were coded as (6=5), (5=4), (4=3), (3=2), (2=1), (1=0). Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Parents perception of school quality (PQSCHOOL)	PA09Q01 PA09Q02 PA09Q03 PA09Q04 PA09Q05 PA09Q06 PA09Q07	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were coded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Parental involvement in their child's school (PARINVOL)	PA10Q01 PA10Q02 PA10Q03 PA10Q04 PA10Q05 PA10Q06 PA10Q07 PA10Q08 PA10Q09 PA10Q10 PA10Q11	1 Yes 2 No Missing	Items were recoded as (1=1), (2=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Which of the following best describes the schooling available to students in your location	PA11Q01	1 Two or more other schools 2 One other school 3 No other schools Missing	All items were coded as (1=2), (2=1), (3=0). Two dummy variable were created as follows: Value value – median 0 Missing 0 1
Parent school selection	PA12Q01 PA12Q02 PA12Q03 PA12Q04 PA12Q05 PA12Q06 PA12Q07 PA12Q08 PA12Q09 PA12Q10 PA12Q11	1 Not important 2 Somewhat important 3 Important 4 Very Important Missing	All items were coded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Parents current support of a child (PARSUPP)	PA13Q01 PA13Q02 PA13Q03 PA13Q04 PA13Q05 PA13Q06 PA13Q07	1 Never or hardly ever 2 Once or twice a year 3 Once or twice a month 4 Once or twice a week 5 Every day or almost Missing	Items were coded as (1=0), (2=1), (3=2), (4=3), (5=4). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Mathematics in child's career and job market	PA14Q01 PA14Q02 PA14Q03 PA14Q04	1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Missing	All items were coded as (1=3), (2=2), (3=1), (4=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Academic and professional expectations in mathematics (PQMCAR)	PA15Q01 PA15Q02 PA15Q03 PA15Q04 PA15Q05	1 Yes 2 No Missing	Items were recoded as (1=1), (2=0). Total score was calculated as a ratio of a sum of all questions over maximum score of valid responses (questions with missing value did not contribute to max score). Two dummy variables were created as follows: Value value – mean 0 Missing 0 1
Child repeated a grade at <ISCED 1>	PA18Q01	1 No, never 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Child repeated a grade at <ISCED 2>	PA18Q02	1 No, never 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Child repeated a grade at <ISCED 3>	PA18Q03	1 No, never 2 Yes, once 3 Yes, twice or more Missing (or invalid)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
What level of education do you expect your child to complete - ISCED 2 - ISCED 3B or C - ISCED 3A - ISCED 4 - ISCED 5B - ISCED 5A, 6 (PQOCCASP)	PA19Q01 PA19Q02 PA19Q03 PA19Q04 PA19Q05 PA19Q06	1 Tick Missing	Item PA19Q01 was recoded as (1=1) Item PA19Q02 was recoded as (1=2) Item PA19Q03 was recoded as (1=3) Item PA19Q04 was recoded as (1=4) Item PA19Q05 was recoded as (1=5) Item PA19Q06 was recoded as (1=6). New variable PQOCCASP was created as maximum value of six items, thus having categories from 1 to 6. Plus one category for missing (when all six items are missing) Six dummy variables were created based on the value of PQOCCASP and with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Expected by parents occupation of the child (SEI)	BSMJ5 (based on PA20Q01)	16-90 (decimal) Missing	value-median 0 0 1

[Part 3/3]

Table B.4 PISA 2012 Main Survey contrast coding used in conditioning for the parent questionnaire variables

Variable	Variable name	Variable coding	Contrast coding
PARENT QUESTIONNAIRE			
Immigration status of mother (PQIMMIGM) by analogy with ST20int	PA21int (PQCTmother) (PQCTMGmot) (PQCTMGfat)	1 Native 2 Second-Generation 3 First-Generation Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Immigration status of father (PQIMMIGF) by analogy with ST20int	PA21int (PQCTfather) (PQCTFGmot) (PQCTFGfat)	1 Native 2 Second-Generation 3 First-Generation Missing	Three dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01' (including missing)
Age of father when arrived to the country	PA22Q01	Value (decimal) Missing	value – median 0 0 1
Age of mother when arrived to the country	PA23Q01	Value (decimal) Missing	value – median 0 0 1
Citizenships of mother (PQCITIZM)	Based on PA24N0101 PA24N0102 PA24N0103 PA24N0104 PA24N0105 PA24N0106 PA24N0107 PA24N0108 PA24N0109 PA24N0110 PA24N0111 PA24N0112 PA24N0113	1 Country of test only 2 Country of test and other 3 Not country of test (inc. Missing)	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01'
Citizenships of father (PQCITIZF)	Based on PA24N0201 PA24N0202 PA24N0203 PA24N0204 PA24N0205 PA24N0206 PA24N0207 PA24N0208 PA24N0209 PA24N0210 PA24N0211 PA24N0212 PA24N0213	1 Country of test only 2 Country of test and other 3 Not country of test Missing	Two dummy variables with default value of '00' and - national mode = '-1' in all dummies - corresponding category = '01'
Language spoken at home – Father (PQLANGNF)	Based on PA25 in F	1 Language of test 2 Other language Missing	Two dummy variables with default value of '00' and - language of test = '-1' in all dummies - other language = '01' '00' - missing = '00' '01'
Language spoken at home – Mother (PQLANGNM)	Based on PA25 in M	1 Language of test 2 Other language Missing	Two dummy variables with default value of '00' and - language of test = '-1' in all dummies - other language = '01' '00' - missing = '00' '01'

Table B.5 PISA 2012 Main Survey contrast coding used in conditioning for other variables

Variable	Variable name	Variable coding	Contrast coding
OTHER VARIABLES			
School identification number	SCHOOLID	Unique 7-digit school ID	IDs for small schools (less than 8 students) were recoded into '9999999' for schools which did not administer UH booklet to students, '9999998' for schools which administered UH booklet to all students and '8888888' for schools which administered both UH and normal booklet to all students Total number of schools minus one dummies were created for school membership with default value of '00' and - largest school in the country = '-1' in all dummies - corresponding SCHOOLID= '01'
Booklet number	BOOKID	1 or 21	01 00 00 00 00 00 00 00 00 00 00 00
		2 or 22	00 01 00 00 00 00 00 00 00 00 00 00
		3 or 23	00 00 01 00 00 00 00 00 00 00 00 00
		4 or 24	00 00 00 01 00 00 00 00 00 00 00 00
		5 or 25	00 00 00 00 01 00 00 00 00 00 00 00
		6 or 26	00 00 00 00 00 01 00 00 00 00 00 00
		7 or 27	00 00 00 00 00 00 01 00 00 00 00 00
		8	00 00 00 00 00 00 00 01 00 00 00 00
		9	00 00 00 00 00 00 00 00 01 00 00 00
		10	00 00 00 00 00 00 00 00 00 01 00 00
		11	00 00 00 00 00 00 00 00 00 00 01 00
		12	00 00 00 00 00 00 00 00 00 00 00 01
		13	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
		20 (UH)	00 00 00 00 00 00 00 00 00 00 00 00
		71	-1 -1 -1 11 11 11 11 11 11 11 11 11
		72	01 00 00 11 11 11 11 11 11 11 11 11
		73	00 01 00 11 11 11 11 11 11 11 11 11
		74	00 00 01 11 11 11 11 11 11 11 11 11
		70 (UH)	00 00 00 11 11 11 11 11 11 11 11 11

ANNEX C – STANDARD ERRORS OF MEANS, SAMPLE SIZES, SCHOOL VARIANCE ESTIMATES, AND OTHER SAMPLING OUTCOMES

[Part 1/2]

Table C.1 **Standard errors of the student performance mean estimate by country, by domain and cycle**

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	3.52	3.49	3.47	2.13	2.15	2.10	2.06	2.24	2.26
	Austria	2.69	2.66	2.75	3.76	3.27	3.44	4.08	3.74	3.92
	Belgium	3.56	3.90	4.29	2.58	2.29	2.48	3.04	2.95	2.48
	Canada	1.56	1.40	1.57	1.75	1.82	2.02	2.44	1.97	2.03
	Chile	3.59	3.68	3.44				4.99	4.58	4.32
	Czech Republic	2.37	2.78	2.43	3.46	3.55	3.38	4.18	3.55	3.48
	Denmark	2.35	2.44	2.81	2.82	2.74	2.97	3.18	2.62	3.11
	Estonia							2.93	2.75	2.52
	Finland	2.58	2.15	2.48	1.64	1.87	1.92	2.15	2.30	2.02
	France	2.73	2.71	3.18	2.68	2.50	2.99	4.06	3.17	3.36
	Germany	2.47	2.52	2.43	3.39	3.32	3.64	4.41	3.87	3.80
	Greece	4.97	5.58	4.89	4.10	3.90	3.82	4.04	2.97	3.23
	Hungary	3.95	4.01	4.17	2.47	2.84	2.77	3.28	2.89	2.68
	Iceland	1.45	2.25	2.17	1.56	1.42	1.47	1.95	1.81	1.64
	Ireland	3.24	2.72	3.18	2.63	2.45	2.69	3.54	2.79	3.19
	Israel	8.47	9.31	9.01				4.58	4.35	3.71
	Italy	2.91	2.93	3.05	3.04	3.08	3.13	2.43	2.28	2.02
	Japan	5.21	5.49	5.48	3.92	4.02	4.14	3.65	3.34	3.37
	Korea	2.42	2.76	2.69	3.09	3.24	3.54	3.81	3.76	3.36
	Luxembourg	1.59	1.99	2.32	1.48	0.97	1.50	1.28	1.07	1.05
	Mexico	3.31	3.36	3.18	4.09	3.64	3.49	3.06	2.93	2.71
	Netherlands	3.35	3.61	4.01	2.85	3.13	3.15	2.92	2.59	2.74
	New Zealand	2.78	3.14	2.40	2.46	2.26	2.35	2.99	2.39	2.69
	Norway	2.80	2.77	2.75	2.78	2.38	2.87	3.18	2.64	3.11
	Poland	4.46	5.48	5.12	2.88	2.50	2.86	2.79	2.44	2.34
	Portugal	4.52	4.08	4.00	3.73	3.40	3.46	3.56	3.07	3.02
	Slovak Republic				3.12	3.35	3.71	3.06	2.82	2.59
	Slovenia							0.99	1.04	1.11
	Spain	2.71	3.12	2.95	2.60	2.41	2.61	2.23	2.33	2.57
	Sweden	2.20	2.46	2.51	2.42	2.56	2.72	3.44	2.41	2.37
	Switzerland	4.25	4.38	4.44	3.28	3.38	3.69	3.06	3.15	3.16
	Turkey				5.79	6.74	5.89	4.21	4.90	3.84
	United Kingdom	2.56	2.50	2.69	2.46	2.43	2.52	2.26	2.14	2.29
	United States	7.05	7.64	7.31	3.22	2.95	3.08		4.02	4.22
Partners	**Albania**	3.29	3.08	2.89						
	Argentina	9.86	9.38	8.56				7.17	6.24	6.08
	Azerbaijan							3.12	2.26	2.75
	Brazil	3.10	3.71	3.26	4.58	4.83	4.35	3.74	2.93	2.79
	Bulgaria	4.89	5.67	4.58				6.91	6.13	6.11
	Colombia							5.08	3.78	3.37
	Costa Rica									
	Croatia							2.81	2.37	2.45
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	2.93	3.26	3.01	3.69	4.54	4.26	2.42	2.67	2.47
	Indonesia	3.99	4.54	3.94	3.38	3.91	3.21	5.92	5.63	5.73
	Jordan							3.27	3.30	2.84
	Kazakhstan									
	Kyrgyzstan							3.48	3.41	2.93
	Latvia	5.27	4.46	5.62	3.67	3.69	3.89	3.73	3.03	2.97
	Liechtenstein	4.12	6.99	7.09	3.58	4.12	4.33	3.91	4.21	4.10
	Lithuania							2.98	2.93	2.76
	Macao-China				2.16	2.89	3.03	1.10	1.30	1.06
	Macedonia	1.93	2.68	2.10						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							1.22	1.37	1.06
	Panama									
	Peru	4.42	4.41	3.98						
	Qatar							1.20	1.02	0.86
	Romania	3.47	4.25	3.37				4.69	4.21	4.20
	Russian Federation	4.16	5.46	4.74	3.94	4.20	4.14	4.32	3.87	3.67
	Serbia				3.56	3.75	3.50	3.46	3.51	3.04
	Shanghai-China									
	Singapore									
	Chinese Taipei							3.38	4.10	3.57
	Tamil Nadu-India									
	Thailand	3.24	3.60	3.06	2.81	3.00	2.70	2.59	2.34	2.14
	Trinidad and Tobago									
	Tunisia				2.81	2.54	2.56	4.02	3.96	2.96
	United Arab Emirates									
	Uruguay				3.43	3.29	2.90	3.43	2.61	2.75
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median	3.10	3.26	3.18	2.88	3.00	3.08	3.18	2.89	2.79
	Mean	3.32	3.61	3.58	3.00	2.99	3.08	3.23	2.92	2.92

---

1. Footnote by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.1 Standard errors of the student performance mean estimate by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	2.34	2.53	2.53	1.58	1.64	1.76
	Austria	2.95	2.66	3.24	2.76	2.67	2.70
	Belgium	2.35	2.25	2.52	2.25	2.14	2.18
	Canada	1.48	1.61	1.62	1.93	1.84	1.93
	Chile	3.13	3.06	2.92	2.90	3.07	2.86
	Czech Republic	2.89	2.83	2.97	2.87	2.85	2.96
	Denmark	2.07	2.60	2.48	2.65	2.29	2.74
	Estonia	2.64	2.57	2.67	2.03	2.02	1.95
	Finland	2.25	2.17	2.34	2.38	1.94	2.20
	France	3.44	3.09	3.60	2.83	2.45	2.58
	Germany	2.66	2.86	2.80	2.82	2.88	2.96
	Greece	4.32	3.88	4.04	3.27	2.50	3.12
	Hungary	3.17	3.45	3.14	3.16	3.19	2.95
	Iceland	1.41	1.39	1.41	1.80	1.70	2.12
	Ireland	2.97	2.54	3.27	2.55	2.25	2.45
	Israel	3.63	3.28	3.11	5.01	4.68	4.96
	Italy	1.57	1.86	1.77	1.97	2.03	1.94
	Japan	3.47	3.33	3.41	3.67	3.59	3.60
	Korea	3.46	4.02	3.44	3.94	4.58	3.66
	Luxembourg	1.25	1.18	1.23	1.54	1.09	1.30
	Mexico	1.95	1.83	1.79	1.51	1.35	1.31
	Netherlands	5.15	4.75	5.42	3.47	3.47	3.51
	New Zealand	2.35	2.31	2.58	2.40	2.21	2.14
	Norway	2.58	2.40	2.60	3.22	2.73	3.09
	Poland	2.60	2.84	2.41	3.14	3.62	3.12
	Portugal	3.07	2.91	2.90	3.75	3.81	3.75
	Slovak Republic	2.54	3.08	2.99	4.17	3.43	3.61
	Slovenia	1.03	1.23	1.15	1.22	1.23	1.29
	Spain	2.02	2.11	2.05	1.91	1.90	1.83
	Sweden	2.88	2.90	2.72	3.00	2.26	3.00
	Switzerland	2.44	3.30	2.82	2.57	3.04	2.71
	Turkey	3.52	4.44	3.60	4.21	4.83	3.89
	United Kingdom	2.28	2.42	2.52	3.50	3.30	3.38
	United States	3.65	3.57	3.64	3.74	3.60	3.78
Partners	Albania	4.04	3.98	3.94	3.20	2.00	2.44
	Argentina	4.63	4.09	4.58	3.70	3.53	3.88
	Azerbaijan	3.33	2.76	3.05			
	Brazil	2.73	2.39	2.43	2.11	2.06	2.14
	Bulgaria	6.68	5.86	5.86	6.02	3.99	4.78
	Colombia	3.74	3.24	3.63	3.45	2.89	3.05
	Costa Rica	3.17	2.98	2.76	3.50	3.04	2.94
	Croatia	2.87	3.09	2.83	3.31	3.54	3.10
	Cyprus[1, 2]				1.18	1.07	1.18
	Georgia	2.89	2.79	2.92			
	Himachal Pradesh-India	3.99	4.19	4.18			
	Hong Kong-China	2.12	2.73	2.75	2.79	3.22	2.61
	Indonesia	3.74	3.72	3.78	4.21	4.04	3.82
	Jordan	3.31	3.71	3.54	3.56	3.12	3.12
	Kazakhstan	3.07	3.04	3.13	2.69	3.03	2.97
	Kyrgyzstan	3.19	2.87	2.92			
	Latvia	2.96	3.07	3.07	2.39	2.75	2.75
	Liechtenstein	2.80	4.06	3.42	4.10	3.95	3.55
	Lithuania	2.39	2.62	2.93	2.48	2.64	2.55
	Macao-China	0.89	0.92	1.03	0.91	0.96	0.85
	Macedonia						
	Malaysia	2.86	2.71	2.68	3.33	3.18	3.00
	Malta	1.58	1.43	1.72			
	Mauritius	1.09	0.97	1.12			
	Miranda-Venezuela	5.28	4.31	4.90			
	Moldova	2.83	3.13	3.01			
	Montenegro	1.72	2.03	2.03	1.18	1.05	1.07
	Panama	6.54	5.25	5.74			
	Peru	3.95	4.00	3.49	4.34	3.69	3.58
	Qatar	0.76	0.70	0.89	0.82	0.76	0.75
	Romania	4.09	3.41	3.36	3.98	3.76	3.25
	Russian Federation	3.34	3.29	3.30	2.97	3.04	2.85
	Serbia	2.43	2.92	2.37	3.44	3.39	3.40
	Shanghai-China	2.40	2.82	2.30	2.86	3.29	3.03
	Singapore	1.06	1.44	1.36	1.37	1.32	1.51
	Chinese Taipei	2.60	3.40	2.63	3.03	3.30	2.33
	Tamil Nadu-India	5.54	5.11	4.22			
	Thailand	2.64	3.23	2.98	3.08	3.45	2.93
	Trinidad and Tobago	1.24	1.28	1.24			
	Tunisia	2.88	2.98	2.69	4.51	3.91	3.46
	United Arab Emirates	2.86	2.46	2.62	2.50	2.43	2.81
	Uruguay	2.60	2.59	2.57	3.16	2.76	2.77
	Viet Nam				4.40	4.84	4.31
	Central tendency indices on 35 countries that participated in the five surveys						
	Median	2.66	2.83	2.80	2.82	2.73	2.75
	Mean	2.72	2.80	2.83	2.79	2.73	2.73

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table C.2 **Sample sizes by country and cycle**

		PISA 2000			PISA 2003			PISA 2006		
		School sample size	Overall student sample size	Average within-school sample size	School sample size	Overall student sample size	Average within-school sample size	School sample size	Overall student sample size	Average within-school sample size
OECD	**Australia**	231	5 176	22.4	321	12 551	39.1	356	14 170	39.8
	Austria	213	4 745	22.3	193	4 597	23.8	199	4 927	24.8
	Belgium	216	6 670	30.9	277	8 796	31.8	269	8 857	32.9
	Canada	1 117	29 687	26.6	1 087	27 953	25.7	896	22 646	25.3
	Chile	179	4 889	27.3				173	5 233	30.2
	Czech Republic	229	5 365	23.4	260	6 320	24.3	245	5 932	24.2
	Denmark	225	4 235	18.8	206	4 218	20.5	211	4 532	21.5
	Estonia							169	4 865	28.8
	Finland	155	4 864	31.4	197	5 796	29.4	155	4 714	30.4
	France	177	4 673	26.4	170	4 300	25.3	182	4 716	25.9
	Germany	219	5 073	23.2	216	4 660	21.6	226	4 891	21.6
	Greece	157	4 672	29.8	171	4 627	27.1	190	4 873	25.6
	Hungary	194	4 887	25.2	253	4 765	18.8	189	4 490	23.8
	Iceland	130	3 372	25.9	129	3 350	26.0	139	3 789	27.3
	Ireland	139	3 854	27.7	145	3 880	26.8	165	4 585	27.8
	Israel	165	4 498	27.3				149	4 584	30.8
	Italy	172	4 984	29.0	406	11 639	28.7	799	21 773	27.3
	Japan	135	5 256	38.9	144	4 707	32.7	185	5 952	32.2
	Korea	146	4 982	34.1	149	5 444	36.5	154	5 176	33.6
	Luxembourg	24	3 528	147.0	29	3 923	135.3	31	4 567	147.3
	Mexico	183	4 600	25.1	1 124	29 983	26.7	1 140	30 971	27.2
	Netherlands	100	2 503	25.0	154	3 992	25.9	185	4 871	26.3
	New Zealand	153	3 667	24.0	173	4 511	26.1	170	4 823	28.4
	Norway	176	4 147	23.6	182	4 064	22.3	203	4 692	23.1
	Poland	127	3 654	28.8	166	4 383	26.4	221	5 547	25.1
	Portugal	149	4 585	30.8	153	4 608	30.1	173	5 109	29.5
	Slovak Republic				281	7 346	26.1	189	4 731	25.0
	Slovenia							361	6 595	18.3
	Spain	185	6 214	33.6	383	10 791	28.2	686	19 604	28.6
	Sweden	154	4 416	28.7	185	4 624	25.0	197	4 443	22.6
	Switzerland	282	6 100	21.6	445	8 420	18.9	510	12 192	23.9
	Turkey				159	4 855	30.5	160	4 942	30.9
	United Kingdom	362	9 340	25.8	383	9 535	24.9	502	13 152	26.2
	United States	153	3 846	25.1	274	5 456	19.9	166	5 611	33.8
Partners	**Albania**	174	4 980	28.6						
	Argentina	156	3 983	25.5				176	4 339	24.7
	Azerbaijan							171	5 184	30.3
	Brazil	324	4 893	15.1	228	4 452	19.5	625	9 295	14.9
	Bulgaria	160	4 657	29.1				180	4 498	25.0
	Colombia							165	4 478	27.1
	Costa Rica									
	Croatia							161	5 213	32.4
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	140	4 405	31.5	145	4 478	30.9	146	4 645	31.8
	Indonesia	290	7 368	25.4	346	10 761	31.1	352	10 647	30.2
	Jordan							210	6 509	31.0
	Kazakhstan									
	Kyrgyzstan							201	5 904	29.4
	Latvia	154	3 893	25.3	157	4 627	29.5	176	4 719	26.8
	Liechtenstein	11	314	28.5	12	332	27.7	12	339	28.3
	Lithuania							197	4 744	24.1
	Macao-China				39	1 250	32.1	43	4 760	110.7
	Macedonia	91	4 510	49.6						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							51	4 455	87.4
	Panama									
	Peru	177	4 429	25.0						
	Qatar							131	6 265	47.8
	Romania	177	4 829	27.3				174	5 118	29.4
	Russian Federation	246	6 701	27.2	212	5 974	28.2	209	5 799	27.7
	Serbia				149	4 405	29.6	162	4 798	29.6
	Shanghai-China									
	Singapore									
	Chinese Taipei							236	8 815	37.4
	Tamil Nadu-India									
	Thailand	179	5 340	29.8	179	5 236	29.3	212	6 192	29.2
	Trinidad and Tobago									
	Tunisia				149	4 721	31.7	152	4 640	30.5
	United Arab Emirates									
	Uruguay				243	5 835	24.0	278	4 839	17.4
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median			26.4			26.7			27.3
	Mean			30.2			29.8			30.7

[1] Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

[2] Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.2 Sample sizes by country and cycle

		PISA 2009			PISA 2012		
		School sample size	Overall student sample size	Average within-school sample size	School sample size	Overall student sample size	Average within-school sample size
OECD	**Australia**	353	14 251	40.4	775	17 774	22.9
	Austria	282	6 590	23.4	191	4 756	24.9
	Belgium	278	8 501	30.6	287	9 690	33.8
	Canada	978	23 207	23.7	885	21 548	24.3
	Chile	200	5 669	28.3	221	6 857	31.0
	Czech Republic	261	6 064	23.2	297	6 535	22.0
	Denmark	285	5 924	20.8	341	7 481	21.9
	Estonia	175	4 727	27.0	206	5 867	28.5
	Finland	203	5 810	28.6	311	8 829	28.4
	France	168	4 298	25.6	226	5 682	25.1
	Germany	226	4 979	22.0	230	5 001	21.7
	Greece	184	4 969	27.0	188	5 125	27.3
	Hungary	187	4 605	24.6	204	4 810	23.6
	Iceland	131	3 646	27.8	134	3 508	26.2
	Ireland	144	3 937	27.3	183	5 016	27.4
	Israel	176	5 761	32.7	172	6 061	35.2
	Italy	1 097	30 905	28.2	1 194	38 142	31.9
	Japan	186	6 088	32.7	191	6 351	33.3
	Korea	157	4 989	31.8	156	5 033	32.3
	Luxembourg	39	4 622	118.5	42	5 260	125.2
	Mexico	1 535	38 250	24.9	1 471	33 806	23.0
	Netherlands	186	4 760	25.6	179	4 460	24.9
	New Zealand	163	4 643	28.5	177	5 248	29.6
	Norway	197	4 660	23.7	197	4 686	23.8
	Poland	185	4 917	26.6	184	5 662	30.8
	Portugal	214	6 298	29.4	195	5 722	29.3
	Slovak Republic	189	4 555	24.1	231	5 737	24.8
	Slovenia	341	6 155	18.0	338	7 229	21.4
	Spain	889	25 887	29.1	902	25 335	28.1
	Sweden	189	4 567	24.2	209	4 739	22.7
	Switzerland	426	11 812	27.7	411	11 234	27.3
	Turkey	170	4 996	29.4	170	4 848	28.5
	United Kingdom	482	12 179	25.3	507	12 659	25.0
	United States	165	5 233	31.7	162	6 111	37.7
Partners	**Albania**	181	4 596	25.4	204	4 743	23.3
	Argentina	199	4 774	24.0	226	5 908	26.1
	Azerbaijan	162	4 691	29.0			
	Brazil	947	20 127	21.3	839	20 091	23.9
	Bulgaria	178	4 507	25.3	188	5 282	28.1
	Colombia	275	7 921	28.8	352	11 173	31.7
	Costa Rica	181	4 578	25.3	193	4 602	23.8
	Croatia	158	4 994	31.6	163	6 153	37.7
	Cyprus[1, 2]				117	5 078	43.4
	Georgia	226	4 646	20.6			
	Himachal Pradesh-India	66	1 616	24.5			
	Hong Kong-China	151	4 837	32.0	148	4 670	31.6
	Indonesia	183	5 136	28.1	209	5 622	26.9
	Jordan	210	6 486	30.9	233	7 038	30.2
	Kazakhstan	199	5 412	27.2	218	5 808	26.6
	Kyrgyzstan	173	4 986	28.8			
	Latvia	184	4 502	24.5	211	5 276	25.0
	Liechtenstein	12	329	27.4	12	293	24.4
	Lithuania	196	4 528	23.1	216	4 618	21.4
	Macao-China	45	5 952	132.3	45	5 335	118.6
	Macedonia						
	Malaysia	152	4 999	32.9	164	5 197	31.7
	Malta	53	3 453	65.2			
	Mauritius	185	4 654	25.2			
	Miranda-Venezuela	121	2 901	24.0			
	Moldova	186	5 194	27.9			
	Montenegro	52	4 825	92.8	51	4 744	93.0
	Panama	188	3 969	21.1			
	Peru	240	5 985	24.9	240	6 035	25.1
	Qatar	153	9 078	59.3	157	10 966	69.8
	Romania	159	4 776	30.0	178	5 074	28.5
	Russian Federation	213	5 308	24.9	227	6 418	28.3
	Serbia	190	5 523	29.1	153	4 684	30.6
	Shanghai-China	152	5 115	33.7	155	6 374	41.1
	Singapore	171	5 283	30.9	172	5 546	32.2
	Chinese Taipei	158	5 831	36.9	163	6 046	37.1
	Tamil Nadu-India	147	3 210	21.8			
	Thailand	230	6 225	27.1	239	6 606	27.6
	Trinidad and Tobago	158	4 778	30.2			
	Tunisia	165	4 955	30.0	153	4 407	28.8
	United Arab Emirates	369	10 867	29.4	458	11 500	25.1
	Uruguay	232	5 957	25.7	180	5 315	29.5
	Viet Nam				162	4 959	30.6
	Central tendency indices on 35 countries that participated in the five surveys						
	Median			27.1			26.9
	Mean			29.7			29.8

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table C.3 School variance estimate by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	1 888	1 405	1 500	2 009	1 927	2 079	1 878	1 694	1 839
	Austria	6 417	5 173	5 241	7 566	5 250	5 823	6 861	5 785	5 464
	Belgium	7 025	6 291	6 939	7 186	7 240	5 983	6 593	5 814	5 182
	Canada	1 588	1 255	1 279	1 199	1 270	1 492	2 163	1 547	1 668
	Chile	4 968	4 208	3 702				6 011	4 800	4 740
	Czech Republic	4 814	4 055	3 612	4 507	4 942	4 388	7 325	6 451	5 617
	Denmark	1 876	1 363	1 760	1 437	1 147	1 308	1 593	1 281	1 393
	Estonia							2 217	1 594	1 437
	Finland	1 009	410	448	257	343	361	643	489	433
	France	4 243	3 704	5 006	4 245	3 830	5 803	6 090	5 049	5 488
	Germany	6 903	5 653	5 191	7 001	6 101	7 036	9 733	6 183	5 944
	Greece	5 060	5 576	3 786	3 976	3 357	2 723	5 493	3 877	4 369
	Hungary	6 408	5 236	5 731	4 919	5 710	5 424	7 164	6 181	5 453
	Iceland	696	430	572	382	319	365	1 220	725	898
	Ireland	1 566	816	1 242	1 712	1 218	1 408	2 010	1 310	1 539
	Israel	5 109	5 673	4 953				5 641	4 668	3 926
	Italy	4 844	3 578	4 188	5 009	4 915	5 701	6 210	4 951	4 758
	Japan	3 377	3 727	3 646	4 998	5 400	5 543	5 459	4 474	4 867
	Korea	1 840	2 889	2 574	2 475	3 607	3 870	3 205	3 494	2 869
	Luxembourg	3 069	2 056	2 474	2 656	2 673	3 018	2 817	2 777	2 738
	Mexico	3 969	3 467	2 429	2 818	2 496	1 934	3 296	2 580	2 293
	Netherlands	3 984	3 873	4 262	4 316	5 508	5 743	5 567	4 880	5 359
	New Zealand	1 892	1 702	1 732	1 916	1 781	1 922	2 108	1 406	1 930
	Norway	1 111	726	845	819	578	846	1 385	942	964
	Poland	6 127	5 483	4 684	1 351	1 035	1 489	1 580	1 121	1 108
	Portugal	3 457	2 492	2 427	3 315	2 620	2 733	3 449	2 746	2 502
	Slovak Republic				3 538	3 794	4 560	5 567	4 541	3 690
	Slovenia							6 634	4 674	5 811
	Spain	1 473	1 445	1 595	1 700	1 489	1 677	1 271	1 240	1 151
	Sweden	793	691	679	873	970	1 046	1 694	1 215	1 091
	Switzerland	4 421	3 970	4 024	2 608	3 165	3 314	3 101	3 283	3 375
	Turkey				4 772	5 915	4 732	4 047	4 557	3 653
	United Kingdom	2 114	1 865	2 195	1 857	1 892	2 089	2 234	1 726	2 200
	United States	3 236	3 127	3 637	2 481	2 345	2 270		2 201	2 626
Partners	**Albania**	4 046	3 355	2 521						
	Argentina	5 920	6 282	4 897				6 881	5 072	4 794
	Azerbaijan							2 359	1 655	1 612
	Brazil	3 379	3 548	2 453	3 416	4 159	3 182	4 555	4 342	3 711
	Bulgaria	6 162	5 732	3 781				7 870	5 199	6 226
	Colombia							3 466	2 973	2 244
	Costa Rica									
	Croatia							3 794	2 721	3 036
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	3 318	3 955	3 198	2 949	4 573	3 915	2 605	3 420	3 072
	Indonesia	2 019	2 253	1 704	1 991	2 720	1 605	2 422	2 746	1 745
	Jordan							2 629	1 660	1 792
	Kazakhstan									
	Kyrgyzstan							4 334	3 159	2 763
	Latvia	3 305	2 836	2 775	1 666	1 761	1 778	2 183	1 537	1 316
	Liechtenstein	3 456	3 395	3 171	2 998	3 461	3 510	3 452	2 921	3 176
	Lithuania							2 671	2 687	2 308
	Macao-China				1 105	1 455	1 356	1 708	1 733	1 739
	Macedonia	3 994	3 019	2 350						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							2 715	1 752	1 812
	Panama									
	Peru	5 992	4 842	2 504						
	Qatar							7 141	5 015	4 240
	Romania	5 139	5 361	3 235				4 658	3 614	3 182
	Russian Federation	3 079	3 896	3 034	2 034	2 558	2 086	3 121	2 325	2 166
	Serbia				2 305	2 566	1 978	3 941	3 723	3 086
	Shanghai-China									
	Singapore									
	Chinese Taipei							3 194	5 020	4 120
	Tamil Nadu-India									
	Thailand	1 848	2 324	1 789	2 120	2 602	2 176	2 863	2 480	2 294
	Trinidad and Tobago									
	Tunisia				3 024	2 807	2 549	4 636	4 003	2 904
	United Arab Emirates									
	Uruguay				5 553	4 618	4 108	6 018	3 926	3 525
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median	3 305	3 127	2 574	2 481	2 620	2 270	2 982	2 746	2 502
	Mean	3 303	2 990	2 909	2 936	2 999	3 018	3 628	3 006	2 931

[1] Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

[2] Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.3 School variance estimate by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	2 102	2 031	2 243	2 648	2 616	2 522
	Austria	5 886	5 143	5 905	4 876	5 046	4 838
	Belgium	6 358	6 769	7 501	6 276	6 004	5 653
	Canada	1 541	1 602	1 492	1 635	1 518	1 500
	Chile	3 862	3 485	3 148	4 138	4 382	4 000
	Czech Republic	5 175	5 596	6 359	4 481	5 536	4 500
	Denmark	1 114	1 319	1 499	1 603	1 252	1 668
	Estonia	1 547	1 399	1 420	1 506	1 261	1 299
	Finland	550	554	597	1 205	935	1 086
	France	6 455	5 599	5 906	7 189	5 577	5 787
	Germany	5 867	6 255	6 659	5 298	5 351	5 281
	Greece	3 812	2 663	3 296	5 099	3 206	3 452
	Hungary	6 303	6 022	5 293	6 036	5 997	4 992
	Iceland	1 380	1 592	1 655	1 412	1 074	1 157
	Ireland	2 256	1 590	2 124	1 778	1 362	1 627
	Israel	6 130	4 919	4 781	6 021	4 757	5 140
	Italy	5 055	4 245	4 582	5 566	4 585	4 527
	Japan	5 093	5 090	4 911	4 615	4 841	4 169
	Korea	2 052	2 989	2 184	2 736	3 831	2 418
	Luxembourg	3 585	3 138	4 095	3 305	2 938	3 322
	Mexico	3 002	2 481	2 266	2 588	2 010	1 871
	Netherlands	4 698	4 911	5 770	5 794	5 655	5 714
	New Zealand	2 200	2 101	2 537	2 904	2 494	2 854
	Norway	874	802	923	1 567	1 229	1 582
	Poland	1 309	1 335	1 105	2 041	2 105	1 692
	Portugal	2 416	2 674	1 982	2 787	2 841	2 561
	Slovak Republic	3 557	4 288	4 541	6 559	4 496	5 396
	Slovenia	5 306	4 834	5 169	5 836	4 903	5 161
	Spain	1 445	1 543	1 415	1 454	1 299	1 203
	Sweden	1 514	1 576	1 594	2 282	1 267	1 824
	Switzerland	2 624	3 158	3 084	2 704	2 867	2 724
	Turkey	4 118	5 876	4 049	4 570	5 126	3 763
	United Kingdom	1 796	1 804	2 177	2 117	1 983	2 259
	United States	2 235	2 458	2 606	2 247	1 971	2 276
Partners	**Albania**	2 856	2 754	2 355	923	451	625
	Argentina	6 532	4 863	5 954	4 775	3 164	4 225
	Azerbaijan	2 533	1 960	2 289			
	Brazil	3 315	2 720	2 795	3 269	2 753	2 752
	Bulgaria	8 333	5 725	6 753	9 527	4 957	6 326
	Colombia	2 706	2 107	2 376	2 721	2 077	2 034
	Costa Rica	2 391	1 944	1 986	2 062	1 810	1 871
	Croatia	3 418	3 014	2 911	3 733	3 296	2 740
	Cyprus[1, 2]				4 871	3 318	3 568
	Georgia	2 184	2 017	1 795			
	Himachal Pradesh-India	1 967	1 986	2 418			
	Hong Kong-China	2 944	3 753	3 073	3 114	4 064	2 719
	Indonesia	2 070	2 364	2 097	2 663	2 333	2 061
	Jordan	2 809	2 594	2 493	3 485	1 977	2 390
	Kazakhstan	3 159	2 909	2 784	2 081	1 974	2 045
	Kyrgyzstan	4 108	2 901	3 302			
	Latvia	1 499	1 553	1 574	2 214	1 792	1 665
	Liechtenstein	2 641	2 212	2 292	3 084	3 820	2 465
	Lithuania	2 360	2 452	2 228	2 968	2 667	2 564
	Macao-China	2 089	1 983	1 804	2 896	3 562	2 385
	Macedonia						
	Malaysia	1 708	1 579	1 539	1 714	2 090	1 615
	Malta	8 366	5 811	7 233			
	Mauritius	6 707	4 966	5 278			
	Miranda-Venezuela	5 058	3 665	4 644			
	Moldova	2 455	2 633	2 422			
	Montenegro	2 833	2 262	2 112	3 440	2 267	2 171
	Panama	5 319	3 621	4 515			
	Peru	5 149	4 166	3 787	4 320	3 272	2 759
	Qatar	7 276	5 374	5 659	6 304	4 679	5 263
	Romania	4 673	2 846	2 920	4 227	3 305	3 091
	Russian Federation	2 224	2 129	2 080	2 657	2 161	2 122
	Serbia	2 914	3 284	2 676	4 044	3 853	3 241
	Shanghai-China	2 830	5 033	2 857	2 988	4 760	3 064
	Singapore	3 239	3 726	3 866	3 496	3 853	3 853
	Chinese Taipei	2 627	4 579	2 751	3 427	5 868	3 063
	Tamil Nadu-India	2 389	1 794	1 462			
	Thailand	2 162	2 769	2 264	2 826	3 758	2 680
	Trinidad and Tobago	8 353	6 489	7 157			
	Tunisia	3 117	2 857	2 746	3 952	3 008	2 675
	United Arab Emirates	5 135	3 979	4 361	4 974	3 852	4 095
	Uruguay	4 153	3 428	3 899	4 268	3 518	3 763
	Viet Nam				3 286	4 063	3 096
	Central tendency indices on 35 countries that participated in the five surveys						
	Median	2 256	2 481	2 266	2 736	2 753	2 522
	Mean	3 016	2 987	3 084	3 259	3 088	2 901

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table C.4 Intraclass correlation by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	0.18	0.17	0.17	0.21	0.21	0.20	0.21	0.22	0.18
	Austria	0.60	0.53	0.55	0.62	0.55	0.57	0.56	0.56	0.55
	Belgium	0.60	0.55	0.55	0.56	0.56	0.50	0.54	0.53	0.52
	Canada	0.18	0.18	0.16	0.15	0.17	0.15	0.23	0.21	0.19
	Chile	0.56	0.45	0.40				0.49	0.56	0.50
	Czech Republic	0.53	0.44	0.41	0.49	0.51	0.42	0.56	0.55	0.53
	Denmark	0.19	0.18	0.16	0.18	0.14	0.13	0.20	0.17	0.16
	Estonia							0.31	0.25	0.21
	Finland	0.12	0.06	0.06	0.04	0.05	0.04	0.10	0.07	0.06
	France	0.50	0.46	0.48	0.45	0.46	0.47	0.57	0.56	0.54
	Germany	0.59	0.55	0.50	0.58	0.58	0.56	0.67	0.61	0.57
	Greece	0.51	0.46	0.40	0.35	0.36	0.27	0.49	0.42	0.47
	Hungary	0.67	0.53	0.53	0.53	0.58	0.51	0.68	0.65	0.61
	Iceland	0.08	0.06	0.07	0.04	0.04	0.04	0.13	0.09	0.09
	Ireland	0.18	0.12	0.15	0.22	0.17	0.16	0.23	0.19	0.17
	Israel	0.43	0.34	0.32				0.38	0.40	0.31
	Italy	0.55	0.42	0.42	0.49	0.52	0.48	0.52	0.52	0.50
	Japan	0.46	0.49	0.44	0.44	0.53	0.46	0.50	0.53	0.47
	Korea	0.37	0.40	0.39	0.36	0.42	0.38	0.40	0.40	0.35
	Luxembourg	0.31	0.24	0.27	0.27	0.31	0.28	0.29	0.32	0.30
	Mexico	0.53	0.50	0.41	0.36	0.39	0.28	0.41	0.42	0.40
	Netherlands	0.50	0.51	0.46	0.58	0.62	0.57	0.62	0.63	0.60
	New Zealand	0.16	0.17	0.17	0.17	0.18	0.18	0.19	0.16	0.17
	Norway	0.10	0.09	0.09	0.08	0.07	0.08	0.13	0.11	0.11
	Poland	0.62	0.55	0.50	0.15	0.13	0.14	0.16	0.15	0.14
	Portugal	0.37	0.30	0.31	0.38	0.34	0.31	0.36	0.33	0.32
	Slovak Republic				0.41	0.43	0.43	0.50	0.49	0.42
	Slovenia							0.73	0.60	0.60
	Spain	0.20	0.18	0.17	0.19	0.20	0.17	0.17	0.16	0.15
	Sweden	0.09	0.08	0.08	0.10	0.11	0.09	0.17	0.15	0.12
	Switzerland	0.43	0.40	0.42	0.30	0.34	0.30	0.37	0.36	0.36
	Turkey				0.53	0.55	0.53	0.48	0.53	0.53
	United Kingdom	0.22	0.23	0.24	0.22	0.23	0.20	0.22	0.23	0.20
	United States	0.29	0.33	0.35	0.24	0.26	0.22		0.28	0.24
Partners	**Albania**	0.41	0.29	0.28						
	Argentina	0.51	0.43	0.41				0.45	0.51	0.48
	Azerbaijan							0.46	0.57	0.50
	Brazil	0.44	0.36	0.30	0.28	0.45	0.34	0.46	0.53	0.47
	Bulgaria	0.56	0.46	0.40				0.56	0.51	0.54
	Colombia							0.30	0.37	0.30
	Costa Rica									
	Croatia							0.47	0.38	0.40
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	0.47	0.45	0.45	0.42	0.47	0.45	0.39	0.40	0.37
	Indonesia	0.43	0.34	0.33	0.36	0.44	0.37	0.50	0.50	0.43
	Jordan							0.31	0.25	0.23
	Kazakhstan									
	Kyrgyzstan							0.41	0.42	0.39
	Latvia	0.31	0.26	0.29	0.20	0.23	0.20	0.26	0.22	0.19
	Liechtenstein	0.45	0.43	0.41	0.43	0.43	0.40	0.46	0.41	0.43
	Lithuania							0.29	0.32	0.28
	Macao-China				0.23	0.19	0.17	0.27	0.23	0.26
	Macedonia	0.45	0.31	0.34						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							0.33	0.25	0.28
	Panama									
	Peru	0.58	0.39	0.30						
	Qatar							0.54	0.53	0.53
	Romania	0.48	0.40	0.35				0.54	0.52	0.49
	Russian Federation	0.37	0.36	0.31	0.23	0.30	0.21	0.35	0.28	0.27
	Serbia				0.34	0.35	0.29	0.45	0.42	0.41
	Shanghai-China									
	Singapore									
	Chinese Taipei							0.46	0.49	0.47
	Tamil Nadu-India									
	Thailand	0.31	0.33	0.30	0.34	0.37	0.32	0.42	0.36	0.37
	Trinidad and Tobago									
	Tunisia				0.33	0.42	0.33	0.47	0.48	0.42
	United Arab Emirates									
	Uruguay				0.36	0.44	0.33	0.41	0.40	0.40
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median	0.37	0.36	0.33	0.30	0.34	0.28	0.38	0.36	0.35
	Mean	0.37	0.34	0.32	0.31	0.33	0.30	0.37	0.35	0.33

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.4 Intraclass correlation by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	0.22	0.23	0.22	0.28	0.28	0.25
	Austria	0.57	0.55	0.54	0.56	0.54	0.53
	Belgium	0.57	0.58	0.60	0.57	0.54	0.52
	Canada	0.19	0.21	0.19	0.19	0.20	0.18
	Chile	0.50	0.50	0.44	0.56	0.55	0.52
	Czech Republic	0.54	0.57	0.57	0.53	0.56	0.51
	Denmark	0.16	0.17	0.17	0.22	0.18	0.19
	Estonia	0.22	0.21	0.20	0.23	0.19	0.20
	Finland	0.07	0.08	0.07	0.13	0.13	0.12
	France	0.58	0.55	0.56	0.59	0.58	0.57
	Germany	0.61	0.61	0.62	0.60	0.55	0.55
	Greece	0.41	0.33	0.38	0.46	0.39	0.41
	Hungary	0.69	0.65	0.64	0.67	0.65	0.59
	Iceland	0.14	0.18	0.17	0.14	0.12	0.12
	Ireland	0.25	0.21	0.22	0.24	0.19	0.19
	Israel	0.48	0.45	0.41	0.46	0.43	0.44
	Italy	0.56	0.50	0.50	0.56	0.53	0.51
	Japan	0.50	0.56	0.49	0.46	0.54	0.45
	Korea	0.33	0.38	0.33	0.36	0.39	0.36
	Luxembourg	0.34	0.33	0.37	0.30	0.32	0.31
	Mexico	0.45	0.43	0.41	0.40	0.36	0.37
	Netherlands	0.62	0.63	0.64	0.66	0.66	0.62
	New Zealand	0.21	0.23	0.22	0.25	0.25	0.26
	Norway	0.10	0.11	0.11	0.15	0.15	0.16
	Poland	0.16	0.17	0.14	0.25	0.25	0.22
	Portugal	0.32	0.32	0.28	0.32	0.32	0.32
	Slovak Republic	0.45	0.46	0.48	0.60	0.46	0.53
	Slovenia	0.64	0.55	0.57	0.64	0.59	0.59
	Spain	0.19	0.19	0.19	0.18	0.17	0.16
	Sweden	0.15	0.18	0.16	0.20	0.15	0.18
	Switzerland	0.32	0.34	0.35	0.35	0.34	0.34
	Turkey	0.55	0.63	0.57	0.58	0.62	0.57
	United Kingdom	0.20	0.25	0.23	0.24	0.24	0.24
	United States	0.24	0.30	0.27	0.26	0.24	0.26
Partners	Albania	0.29	0.34	0.31	0.07	0.05	0.06
	Argentina	0.54	0.54	0.55	0.48	0.49	0.52
	Azerbaijan	0.42	0.43	0.39			
	Brazil	0.41	0.46	0.43	0.43	0.45	0.44
	Bulgaria	0.58	0.54	0.55	0.63	0.55	0.57
	Colombia	0.36	0.36	0.36	0.39	0.37	0.35
	Costa Rica	0.38	0.39	0.36	0.39	0.40	0.39
	Croatia	0.44	0.39	0.40	0.51	0.43	0.38
	Cyprus[1, 2]				0.37	0.36	0.36
	Georgia	0.23	0.28	0.22			
	Himachal Pradesh-India	0.32	0.37	0.42			
	Hong Kong-China	0.41	0.41	0.40	0.42	0.43	0.38
	Indonesia	0.48	0.48	0.45	0.49	0.49	0.46
	Jordan	0.34	0.38	0.31	0.41	0.34	0.35
	Kazakhstan	0.39	0.42	0.38	0.38	0.38	0.38
	Kyrgyzstan	0.41	0.43	0.39			
	Latvia	0.23	0.24	0.25	0.30	0.27	0.27
	Liechtenstein	0.43	0.31	0.34	0.48	0.52	0.43
	Lithuania	0.32	0.32	0.31	0.38	0.33	0.34
	Macao-China	0.33	0.26	0.29	0.38	0.36	0.35
	Macedonia						
	Malaysia	0.26	0.29	0.26	0.25	0.32	0.26
	Malta	0.52	0.47	0.47			
	Mauritius	0.66	0.60	0.60			
	Miranda-Venezuela	0.50	0.52	0.49			
	Moldova	0.31	0.36	0.32			
	Montenegro	0.34	0.31	0.28	0.39	0.34	0.32
	Panama	0.56	0.55	0.55			
	Peru	0.53	0.51	0.47	0.49	0.46	0.45
	Qatar	0.55	0.54	0.53	0.48	0.47	0.46
	Romania	0.57	0.46	0.47	0.50	0.48	0.49
	Russian Federation	0.28	0.29	0.25	0.33	0.29	0.30
	Serbia	0.41	0.39	0.37	0.47	0.47	0.42
	Shanghai-China	0.44	0.47	0.43	0.47	0.47	0.46
	Singapore	0.34	0.35	0.36	0.35	0.35	0.36
	Chinese Taipei	0.35	0.41	0.36	0.40	0.43	0.43
	Tamil Nadu-India	0.42	0.43	0.38			
	Thailand	0.41	0.42	0.35	0.44	0.49	0.43
	Trinidad and Tobago	0.62	0.65	0.61			
	Tunisia	0.43	0.47	0.41	0.50	0.49	0.43
	United Arab Emirates	0.49	0.46	0.46	0.52	0.47	0.45
	Uruguay	0.41	0.40	0.40	0.46	0.44	0.41
	Viet Nam				0.57	0.54	0.50
	Central tendency indices on 35 countries that participated in the five surveys						
	Median	0.33	0.33	0.34	0.35	0.34	0.34
	Mean	0.35	0.36	0.35	0.37	0.36	0.35

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table C.5 Within explicit strata intraclass correlation by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	0.17	0.17	0.16	0.15	0.15	0.14	0.13	0.15	0.11
	Austria	0.12	0.15	0.15	0.42	0.33	0.34	0.31	0.29	0.31
	Belgium	0.42	0.39	0.38	0.31	0.29	0.26	0.33	0.32	0.32
	Canada	0.18	0.18	0.16	0.12	0.13	0.12	0.20	0.18	0.16
	Chile	0.33	0.26	0.21				0.33	0.38	0.32
	Czech Republic	0.12	0.15	0.10	0.32	0.33	0.25	0.29	0.25	0.24
	Denmark	0.18	0.16	0.16	0.17	0.13	0.12	0.19	0.17	0.16
	Estonia							0.21	0.18	0.14
	Finland	0.11	0.06	0.04	0.03	0.04	0.04	0.08	0.07	0.05
	France	0.19	0.17	0.16	0.17	0.16	0.17	0.31	0.26	0.25
	Germany				0.50	0.50	0.48	0.55	0.54	0.49
	Greece	0.43	0.35	0.33	0.33	0.35	0.25	0.39	0.29	0.33
	Hungary	0.59	0.49	0.46	0.20	0.26	0.17	0.43	0.40	0.33
	Iceland	0.07	0.05	0.07	0.03	0.03	0.03	0.11	0.08	0.08
	Ireland	0.17	0.11	0.14	0.20	0.15	0.14	0.21	0.17	0.15
	Israel	0.37	0.28	0.28				0.31	0.30	0.25
	Italy	0.34	0.26	0.27	0.19	0.22	0.18	0.20	0.21	0.17
	Japan	0.44	0.47	0.42	0.43	0.51	0.44	0.46	0.50	0.44
	Korea	0.17	0.13	0.13	0.18	0.21	0.20	0.27	0.25	0.20
	Luxembourg	0.31	0.22	0.27	0.25	0.28	0.25	0.13	0.15	0.14
	Mexico	0.48	0.44	0.36	0.30	0.33	0.23	0.29	0.31	0.29
	Netherlands	0.18	0.18	0.15	0.28	0.30	0.22	0.37	0.30	0.26
	New Zealand	0.15	0.16	0.16	0.17	0.17	0.17	0.19	0.16	0.16
	Norway	0.09	0.08	0.09	0.08	0.07	0.08	0.12	0.10	0.10
	Poland	0.25	0.23	0.20	0.14	0.12	0.13	0.14	0.13	0.12
	Portugal	0.35	0.29	0.29	0.34	0.30	0.27	0.19	0.16	0.14
	Slovak Republic				0.36	0.38	0.38	0.37	0.36	0.27
	Slovenia							0.36	0.26	0.23
	Spain	0.13	0.11	0.10	0.12	0.12	0.11	0.11	0.09	0.09
	Sweden	0.07	0.06	0.06	0.08	0.09	0.08	0.14	0.12	0.10
	Switzerland	0.35	0.32	0.34	0.25	0.29	0.25	0.28	0.27	0.27
	Turkey				0.36	0.40	0.39	0.41	0.49	0.49
	United Kingdom	0.21	0.22	0.23	0.21	0.21	0.19	0.21	0.21	0.19
	United States				0.22	0.24	0.20		0.28	0.24
Partners	**Albania**	0.26	0.19	0.19						
	Argentina	0.40	0.33	0.31				0.37	0.43	0.40
	Azerbaijan							0.37	0.53	0.42
	Brazil	0.43	0.36	0.29	0.17	0.29	0.20	0.41	0.47	0.42
	Bulgaria	0.47	0.37	0.30				0.48	0.43	0.44
	Colombia							0.29	0.36	0.30
	Costa Rica									
	Croatia							0.22	0.17	0.17
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	0.47	0.44	0.44	0.42	0.46	0.45	0.38	0.39	0.36
	Indonesia	0.38	0.28	0.29	0.33	0.40	0.33	0.46	0.44	0.38
	Jordan							0.26	0.21	0.19
	Kazakhstan									
	Kyrgyzstan							0.23	0.25	0.22
	Latvia	0.26	0.23	0.26	0.18	0.20	0.19	0.24	0.19	0.16
	Liechtenstein	0.45	0.43	0.40						
	Lithuania							0.17	0.19	0.16
	Macao-China				0.22	0.16	0.16	0.19	0.17	0.19
	Macedonia	0.31	0.19	0.19						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							0.27	0.23	0.24
	Panama									
	Peru	0.49	0.30	0.23						
	Qatar							0.20	0.20	0.20
	Romania	0.21	0.17	0.15				0.35	0.35	0.31
	Russian Federation	0.31	0.29	0.25	0.15	0.20	0.12	0.26	0.20	0.19
	Serbia				0.33	0.34	0.27	0.40	0.36	0.36
	Shanghai-China									
	Singapore									
	Chinese Taipei							0.37	0.40	0.38
	Tamil Nadu-India									
	Thailand	0.26	0.29	0.24	0.28	0.32	0.26	0.31	0.29	0.27
	Trinidad and Tobago									
	Tunisia				0.33	0.42	0.33	0.18	0.19	0.13
	United Arab Emirates									
	Uruguay				0.22	0.31	0.22	0.25	0.22	0.21
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median	0.25	0.22	0.23	0.20	0.23	0.19	0.26	0.23	0.20
	Mean	0.27	0.24	0.23	0.23	0.24	0.21	0.26	0.25	0.23

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.5 Within explicit strata intraclass correlation by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	0.15	0.18	0.16	0.22	0.23	0.20
	Austria	0.18	0.18	0.18	0.19	0.18	0.17
	Belgium	0.35	0.36	0.34	0.26	0.23	0.24
	Canada	0.16	0.18	0.16	0.16	0.16	0.15
	Chile	0.26	0.29	0.22	0.28	0.25	0.25
	Czech Republic	0.17	0.19	0.17	0.17	0.15	0.17
	Denmark	0.13	0.14	0.14	0.20	0.16	0.17
	Estonia	0.19	0.17	0.16	0.20	0.16	0.16
	Finland	0.06	0.07	0.07	0.12	0.12	0.11
	France	0.29	0.27	0.30	0.27	0.24	0.25
	Germany	0.53	0.55	0.53	0.53	0.48	0.48
	Greece	0.35	0.29	0.33	0.43	0.35	0.38
	Hungary	0.32	0.34	0.32	0.34	0.37	0.31
	Iceland	0.13	0.17	0.15	0.08	0.07	0.06
	Ireland	0.19	0.17	0.17	0.19	0.16	0.15
	Israel	0.32	0.29	0.27	0.27	0.26	0.27
	Italy	0.22	0.26	0.23	0.24	0.25	0.23
	Japan	0.48	0.55	0.47	0.43	0.52	0.43
	Korea	0.19	0.22	0.18	0.28	0.29	0.27
	Luxembourg	0.20	0.20	0.22	0.15	0.16	0.16
	Mexico	0.35	0.34	0.32	0.29	0.27	0.29
	Netherlands	0.36	0.35	0.40	0.40	0.35	0.33
	New Zealand	0.20	0.22	0.21	0.24	0.23	0.24
	Norway	0.10	0.10	0.11	0.15	0.15	0.16
	Poland	0.14	0.14	0.12	0.22	0.21	0.19
	Portugal	0.26	0.26	0.24	0.30	0.30	0.30
	Slovak Republic	0.26	0.29	0.33	0.44	0.27	0.35
	Slovenia	0.20	0.19	0.19	0.19	0.18	0.18
	Spain	0.13	0.11	0.12	0.11	0.10	0.11
	Sweden	0.12	0.15	0.13	0.17	0.12	0.16
	Switzerland	0.25	0.27	0.27	0.25	0.24	0.24
	Turkey	0.32	0.44	0.37	0.40	0.46	0.40
	United Kingdom	0.12	0.15	0.14	0.15	0.13	0.15
	United States	0.20	0.27	0.24	0.23	0.22	0.22
Partners	**Albania**	0.21	0.27	0.26	0.06	0.04	0.05
	Argentina	0.52	0.52	0.53	0.46	0.47	0.51
	Azerbaijan	0.29	0.35	0.26			
	Brazil	0.26	0.33	0.29	0.22	0.22	0.23
	Bulgaria	0.45	0.40	0.43	0.61	0.53	0.55
	Colombia	0.33	0.33	0.33	0.38	0.36	0.33
	Costa Rica	0.27	0.28	0.24	0.30	0.31	0.29
	Croatia	0.23	0.21	0.21	0.29	0.23	0.20
	Cyprus[1, 2]				0.23	0.20	0.21
	Georgia	0.17	0.22	0.18			
	Himachal Pradesh-India	0.24	0.27	0.31			
	Hong Kong-China	0.40	0.40	0.39	0.42	0.42	0.37
	Indonesia						
	Jordan	0.32	0.35	0.29	0.36	0.26	0.28
	Kazakhstan	0.16	0.21	0.17	0.19	0.22	0.20
	Kyrgyzstan	0.21	0.24	0.22			
	Latvia	0.19	0.21	0.23	0.23	0.21	0.24
	Liechtenstein						
	Lithuania	0.16	0.16	0.17	0.21	0.18	0.20
	Macao-China	0.27	0.22	0.23	0.32	0.31	0.29
	Macedonia						
	Malaysia	0.20	0.22	0.18	0.21	0.24	0.21
	Malta	0.20	0.18	0.16			
	Mauritius	0.43	0.41	0.44			
	Miranda-Venezuela	0.31	0.37	0.31			
	Moldova	0.19	0.27	0.25			
	Montenegro	0.20	0.15	0.14	0.18	0.15	0.17
	Panama	0.40	0.39	0.42			
	Peru	0.40	0.37	0.33	0.39	0.35	0.36
	Qatar	0.37	0.32	0.32	0.32	0.24	0.26
	Romania	0.46	0.38	0.39	0.50	0.48	0.48
	Russian Federation	0.19	0.21	0.19	0.23	0.21	0.21
	Serbia	0.13	0.15	0.15	0.26	0.26	0.24
	Shanghai-China	0.24	0.24	0.24	0.25	0.22	0.23
	Singapore	0.34	0.34	0.36	0.31	0.32	0.32
	Chinese Taipei	0.12	0.14	0.13	0.17	0.14	0.15
	Tamil Nadu-India	0.35	0.36	0.30			
	Thailand	0.29	0.33	0.26	0.31	0.36	0.29
	Trinidad and Tobago	0.44	0.48	0.42			
	Tunisia	0.14	0.23	0.16	0.48	0.45	0.40
	United Arab Emirates	0.35	0.28	0.31	0.41	0.29	0.32
	Uruguay	0.21	0.23	0.20	0.23	0.19	0.17
	Viet Nam				0.53	0.49	0.46
	Central tendency indices on 35 countries that participated in the five surveys						
	Median	0.20	0.22	0.22	0.23	0.22	0.23
	Mean	0.23	0.25	0.24	0.25	0.24	0.23

---

1. Footnote by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 1/2]

Table C.6 Percentage of school variance explained by explicit stratification variables by country, by domain and cycle

		PISA 2000			PISA 2003			PISA 2006		
		Reading	Mathematics	Science	Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	**Australia**	33.0	35.0	35.0	36.1	35.4	37.1	43.3	38.1	42.8
	Austria	90.4	84.5	85.8	55.4	59.5	59.8	64.8	68.5	62.0
	Belgium	51.2	48.1	50.2	64.7	68.2	63.9	57.8	56.8	56.1
	Canada	28.0	25.0	24.0	26.2	23.9	22.7	17.7	18.2	19.5
	Chile	60.1	58.2	59.2				50.3	52.0	53.5
	Czech Republic	88.0	78.3	84.9	49.9	53.0	54.7	67.7	71.9	71.3
	Denmark	7.4	11.8	2.9	6.9	7.2	5.2	1.6	1.4	2.0
	Estonia							39.0	31.9	38.9
	Finland	11.4	12.0	28.5	17.6	10.9	19.1	15.4	11.3	11.1
	France	77.4	76.3	79.3	75.9	76.7	77.0	67.4	72.6	71.0
	Germany				28.6	26.2	28.1	41.2	23.6	26.5
	Greece	28.4	36.1	27.0	7.4	7.9	8.6	32.7	41.9	43.2
	Hungary	29.2	17.9	24.0	78.2	75.3	80.5	64.5	64.9	68.1
	Iceland	14.4	13.5	10.8	23.0	28.1	19.7	11.3	12.8	17.1
	Ireland	6.2	4.8	4.3	13.4	10.9	12.4	11.5	14.4	17.9
	Israel	23.7	24.5	18.1				28.1	34.8	25.8
	Italy	58.3	51.5	50.4	74.9	74.0	75.8	77.5	75.8	79.7
	Japan	9.6	9.0	10.6	6.7	7.9	8.2	16.4	13.4	14.2
	Korea	65.8	76.9	75.6	59.1	62.3	60.5	44.4	51.0	52.4
	Luxembourg	1.3	9.7	2.1	8.1	17.0	15.4	62.3	62.4	60.7
	Mexico	20.1	20.0	17.6	23.6	23.2	23.8	41.1	39.4	39.4
	Netherlands	78.7	78.3	79.3	71.9	74.1	79.2	64.7	75.1	76.6
	New Zealand	5.9	7.1	5.6	2.8	6.7	4.8	1.0	3.0	2.0
	Norway	8.8	9.1	6.5	3.4	2.9	1.3	5.6	8.4	7.3
	Poland	80.0	75.4	75.1	5.7	6.3	6.6	12.6	17.9	17.0
	Portugal	8.0	7.7	8.5	16.1	16.6	15.6	57.2	61.7	64.3
	Slovak Republic				19.2	20.6	19.5	41.2	40.1	49.1
	Slovenia							78.8	76.9	80.2
	Spain	43.8	43.7	47.0	44.7	43.1	40.8	41.7	48.3	43.6
	Sweden	27.0	24.1	31.0	19.0	18.3	17.4	23.5	19.0	17.5
	Switzerland	30.3	29.9	28.2	22.0	22.9	21.4	34.1	34.6	33.3
	Turkey				49.5	44.3	42.8	24.0	17.3	17.1
	United Kingdom	4.4	4.0	3.8	7.1	9.1	6.1	6.7	8.4	7.9
	United States				11.3	10.3	11.6		0.7	0.1
Partners	**Albania**	48.1	43.2	40.9						
	Argentina	34.3	36.2	35.0				28.5	26.3	26.6
	Azerbaijan							29.9	15.9	27.0
	Brazil	4.2	3.1	2.7	47.9	48.1	51.1	16.0	21.1	19.0
	Bulgaria	31.5	32.6	33.5				27.2	26.4	31.5
	Colombia							3.8	3.9	3.2
	Costa Rica									
	Croatia							68.7	67.5	69.6
	Cyprus[1, 2]									
	Georgia									
	Himachal Pradesh-India									
	Hong Kong-China	3.3	3.4	3.2	0.7	1.2	0.9	4.2	3.9	4.4
	Indonesia	18.5	23.9	18.8	15.1	18.3	15.8	17.2	20.5	19.3
	Jordan							18.1	18.9	21.1
	Kazakhstan									
	Kyrgyzstan							57.7	55.2	55.6
	Latvia	22.9	19.1	16.5	13.2	14.2	11.3	13.0	19.9	17.2
	Liechtenstein									
	Lithuania							50.6	51.9	51.7
	Macao-China				7.8	13.6	9.6	35.7	28.6	35.2
	Macedonia	44.8	47.8	54.3						
	Malaysia									
	Malta									
	Mauritius									
	Miranda-Venezuela									
	Moldova									
	Montenegro							24.3	9.6	22.3
	Panama									
	Peru	30.3	32.5	30.1						
	Qatar							79.4	77.4	77.3
	Romania	72.0	68.5	67.7				52.9	49.3	53.3
	Russian Federation	23.2	25.7	24.5	41.9	43.7	46.4	33.7	35.5	35.9
	Serbia				8.3	6.8	8.4	19.3	21.7	21.7
	Shanghai-China									
	Singapore									
	Chinese Taipei							30.8	31.9	29.5
	Tamil Nadu-India									
	Thailand	21.6	17.7	25.1	24.7	20.3	26.8	36.6	27.6	36.2
	Trinidad and Tobago									
	Tunisia				2.3	1.9	0.8	75.3	74.4	79.7
	United Arab Emirates									
	Uruguay				48.4	43.4	44.2	53.0	58.0	60.0
	Viet Nam									
	Central tendency indices on 35 countries that participated in the five surveys									
	Median	20.1	17.9	18.8	22.5	21.6	20.5	33.7	25.6	29.9
	Mean	28.7	28.2	28.6	29.5	30.1	30.3	33.5	33.6	34.0

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

[Part 2/2]

Table C.6 Percentage of school variance explained by explicit stratification variables by country, by domain and cycle

		PISA 2009			PISA 2012		
		Reading	Mathematics	Science	Reading	Mathematics	Science
OECD	Australia	33.9	27.6	30.8	27.6	25.9	24.9
	Austria	83.4	82.0	81.5	81.9	81.7	81.9
	Belgium	58.7	59.3	65.8	72.9	74.4	71.9
	Canada	17.0	15.8	15.1	20.5	20.9	23.2
	Chile	64.0	60.0	64.1	69.6	72.1	68.5
	Czech Republic	82.2	82.7	84.0	82.8	86.3	81.4
	Denmark	22.2	24.7	25.8	9.0	15.5	15.5
	Estonia	17.7	22.3	23.4	18.6	20.2	21.9
	Finland	16.8	11.4	12.4	10.3	9.1	15.5
	France	71.0	70.1	66.1	74.5	76.1	73.8
	Germany	25.7	22.2	29.5	25.0	24.5	26.4
	Greece	21.8	16.2	18.8	11.5	14.1	10.4
	Hungary	78.2	73.3	74.5	74.8	68.4	68.6
	Iceland	13.3	10.1	14.7	46.3	50.2	49.1
	Ireland	27.1	26.2	28.3	25.3	21.2	26.8
	Israel	47.7	50.0	47.2	55.6	53.7	52.1
	Italy	77.1	65.1	70.9	75.3	69.4	72.2
	Japan	7.5	6.4	6.4	10.6	9.4	7.8
	Korea	53.9	52.4	53.1	30.5	36.3	34.4
	Luxembourg	50.0	50.3	50.5	57.6	59.6	56.8
	Mexico	36.6	31.4	32.0	37.8	36.5	32.6
	Netherlands	66.2	67.7	61.7	66.3	73.0	69.6
	New Zealand	2.6	3.9	3.1	6.5	8.4	7.4
	Norway	3.2	5.3	4.9	0.2	0.7	0.3
	Poland	17.9	21.3	19.5	16.4	21.4	18.4
	Portugal	23.1	23.0	20.5	8.7	7.7	8.6
	Slovak Republic	57.4	51.9	47.3	48.1	57.0	52.0
	Slovenia	85.6	80.5	81.9	86.9	84.8	84.7
	Spain	39.8	44.5	39.7	41.3	46.8	38.7
	Sweden	23.4	20.6	20.1	14.9	22.5	15.4
	Switzerland	31.2	28.5	31.1	36.3	36.4	37.2
	Turkey	62.0	52.8	56.5	52.6	46.2	50.1
	United Kingdom	47.7	47.1	45.7	47.3	53.5	47.2
	United States	19.5	14.1	16.7	13.7	13.4	16.3
Partners	Albania	36.0	27.2	22.3	20.0	29.7	19.7
	Argentina	7.6	8.2	6.9	6.8	8.1	3.4
	Azerbaijan	43.3	29.1	45.0			
	Brazil	47.8	41.9	46.2	62.0	66.8	62.4
	Bulgaria	40.4	42.9	38.3	7.3	7.5	8.5
	Colombia	10.8	10.9	10.2	7.2	6.2	6.8
	Costa Rica	40.4	38.9	43.9	32.7	34.2	35.9
	Croatia	63.0	58.0	59.6	59.7	60.4	57.9
	Cyprus[1, 2]				51.3	57.0	52.4
	Georgia	29.4	24.5	20.0			
	Himachal Pradesh-India	34.8	37.4	39.3			
	Hong Kong-China	4.7	4.1	5.4	2.8	4.5	5.3
	Indonesia						
	Jordan	7.0	11.2	11.1	19.7	33.1	26.4
	Kazakhstan	68.9	62.7	65.0	61.4	54.2	59.2
	Kyrgyzstan	60.4	58.1	55.4			
	Latvia	20.6	19.3	11.9	31.9	26.5	14.6
	Liechtenstein						
	Lithuania	57.8	58.3	53.1	56.4	53.9	51.6
	Macao-China	24.7	17.5	27.9	23.9	20.2	23.7
	Macedonia						
	Malaysia	25.6	30.6	36.2	20.6	32.1	24.5
	Malta	76.4	75.5	79.2			
	Mauritius	61.6	52.2	48.7			
	Miranda-Venezuela	55.3	46.7	52.6			
	Moldova	46.3	36.2	30.8			
	Montenegro	52.1	60.4	58.1	66.1	65.4	55.9
	Panama	47.7	46.7	41.1			
	Peru	40.8	44.3	44.8	33.2	36.2	32.1
	Qatar	53.1	60.2	57.6	49.3	65.4	58.1
	Romania	36.7	27.1	29.7	2.1	0.6	0.8
	Russian Federation	37.8	34.6	31.0	39.7	33.6	35.8
	Serbia	78.8	71.9	71.4	59.5	58.8	56.1
	Shanghai-China	58.4	63.6	56.2	61.9	67.6	63.8
	Singapore	0.4	0.7	0.5	17.8	15.9	16.2
	Chinese Taipei	75.5	76.6	74.9	68.5	79.2	77.0
	Tamil Nadu-India	27.1	26.4	31.2			
	Thailand	41.7	31.9	36.5	44.1	43.5	43.8
	Trinidad and Tobago	51.0	50.8	51.9			
	Tunisia	77.3	66.0	72.2	11.1	15.2	10.8
	United Arab Emirates	44.7	54.3	45.9	35.6	52.7	41.8
	Uruguay	62.7	54.9	61.9	64.8	69.0	69.8
	Viet Nam				14.3	17.2	13.8
	Central tendency indices on 35 countries that participated in the five surveys						
	Median	31.2	27.6	30.8	31.9	33.6	32.6
	Mean	36.5	34.4	35.0	36.6	37.5	36.2

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

ANNEX D – MAPPING OF ISCED TO YEARS

Table D.1 **Mapping of ISCED to years**

		Completed ISCED level 1 (primary education)	Completed ISCED level 2 (lower secondary education)	Completed ISCED levels 3B or 3C (upper secondary education providing direct access to the labour market or to ISCED 5B programmes)	Completed ISCED level 3A (upper secondary education providing access to ISCED 5A and 5B programmes) and/or ISCED level 4 (non-tertiary post-secondary)	Completed ISCED level 5A (university level tertiary education) or ISCED level 6 (advanced research programmes)	Completed ISCED level 5B (non-university tertiary education)
OECD	**Australia**	6	10	11	12	15	14
	Austria	4	9	12	13	17	15
	Belgium[1]	6	9	12	12	17	15
	Canada	6	9	12	12	17	15
	Chile	6	8	12	12	17	16
	Czech Republic	5	9	11	13	16	16
	Denmark	7	10	13	13	18	16
	Estonia	6	9	12	12	16	15
	Finland	6	9	12	12	16.5	14.5
	France	5	9	12	12	15	14
	Germany	4	10	13	13	18	15
	Greece	6	9	11.5	12	17	15
	Hungary	4	8	10.5	12	16.5	13.5
	Iceland	7	10	13	14	18	16
	Ireland	6	9	12	12	16	14
	Israel	6	9	12	12	15	15
	Italy	5	8	12	13	17	16
	Japan	6	9	12	12	16	14
	Korea	6	9	12	12	16	14
	Luxembourg	6	9	12	13	17	16
	Mexico	6	9	12	12	16	14
	Netherlands	6	10	13	12	16	15
	New Zealand	5.5	10	11	12	15	14
	Norway	6	9	12	12	16	14
	Poland	a	8	11	12	16	15
	Portugal	6	9	12	12	17	15
	Slovak Republic[2]	4	9	12	13	18	16
	Slovenia	4	8	11	12	16	15
	Spain	5	8	10	12	16.5	13
	Sweden	6	9	11.5	12	16	14
	Switzerland	6	9	12.5	12.5	17.5	14.5
	Turkey	5	8	11	11	15	13
	United Kingdom (exclud. Scotland)	6	9	12	13	16	15
	United Kingdom (Scotland)	7	9	11	13	17	15
	United States	6	9	a	12	16	14
Partners	**Albania**	6	9	12	12	16	16
	Argentina	6	10	12	12	17	14.5
	Azerbaijan	4	9	11	11	17	14
	Brazil	4	8	11	11	16	14.5
	Bulgaria	4	8	10	12	17.5	15
	Colombia	5	9	11	11	15.5	14
	Costa Rica	6	9	11	12	14	16
	Croatia	4	8	11	12	17	15
	Hong Kong-China	6	9	11	13	16	14
	Indonesia	6	9	12	12	15	14
	Jordan	6	10	12	12	16	14.5
	Kazakhstan	4	9	11.5	12.5	15	14
	Latvia	4	8	11	11	16	14
	Liechtenstein	5	9	11	13	17	14
	Lithuania	3	8	11	11	16	15
	Macao-China	6	9	11	12	16	15
	Malaysia	6	9	11	13	15	16
	Montenegro	4	8	11	12	16	15
	Peru	6	9	11	11	17	14
	Qatar	6	9	12	12	16	15
	Romania	4	8	11.5	12.5	16	14
	Russian Federation	4	9	11.5	12	15	a
	Serbia	4	8	11	12	17	14.5
	Shanghai-China	6	9	12	12	16	15
	Singapore	6	8	10	11	16	13
	Chinese Taipei	6	9	12	12	16	14
	Thailand	6	9	12	12	16	14
	Tunisia	6	9	12	13	17	16
	United Arab Emirates	5	9	12	12	16	15
	Uruguay	6	9	12	12	17	15
	Viet Nam	5	9	12	12	17	a

1. In Belgium, the distinction between universities and other tertiary schools doesn’t match the distinction between ISCED 5A and ISCED 5B.

2. In the Slovak Republic, university education (ISCED 5A) usually lasts five years and doctoral studies (ISCED 6) lasts three more years. Therefore, university graduates will have completed 18 years of study and graduates of doctoral programmes will have completed 21 years of study.

ANNEX E – NATIONAL HOUSEHOLD POSSESSION ITEMS

[Part 1/2]

Table E.1 **National household possession items**

		ST26Q15	ST26Q16	ST26Q17
OECD	**Australia**	iPad® or other tablet device	A home gym and/or gym membership	Espresso machine
	Austria	A laptop/notebook of your own	Electronic devices for playing (Playstation®, Nintendo®, X-Box®, Wii®)	A digital video camera
	Belgium (Flemish Community)	A plasma, LCD or LED television	An alarm system	Home cinema
	Belgium (French and German Communities)	A home cinema set (LCD or LED screen with home cinema system)	An alarm system	A housekeeper
	Canada	iPod®/An MP3 player	A subscription to a daily newspaper	Air conditioning
	Chile	Cable TV	A digital video camera	Microwave oven
	Czech Republic	Your own notebook (laptop)	Camcorder	A home cinema set (screen, DVD, player, speakers)
	Denmark	A music instrument (e.g. piano, guitar, violin)	Flat screen TV	N/A
	Estonia	Video camera	Digital camera	Plasma or LCD TV
	Finland	A laptop	Flat screen TV	Home alarm system
	France	A flat screen TV	A digital camera (not installed in a mobile phone)	A laptop/portable PC
	Germany	Electronical devices for playing (Playstation®, Nintendo®, X-Box®, Wii®)	A TV in your own room	Audiobooks
	Greece	Home cinema	Garage or parking space	Alarm system
	Hungary	Video games console (e.g. Playstation®)	Tablet computer (e.g. iPad®, Samsung Galaxy Tab®, BlackBerry®)	Digital camera (not part of a phone)
	Iceland	Security watch or system	Satellite dish	Flat screen or projector
	Ireland	A flat-screen television	A bedroom with an en-suite bathroom	A premium cable TV package (e.g. Sky Movies, Sky Sports)
	Israel	4x4 vehicle	Espresso machine	Home cinema system
	Italy	Antique furniture	Alarm system	Air-conditioning
	Japan	Digital camera	Plasma TV/LCD TV	Clothing Dryer
	Korea	Air conditioner	Digital TV (e.g. PDP, LCD, LED)	Kimchi refrigerator (for maturing) --> Air Cleaner machine
	Luxembourg	iPhone® 3 or 4	iPad®	Playstation 3® or Wii®
	Mexico	Cable TV (Sky, Cablevisión, etc.)	Phone line	Microwave oven
	Netherlands	An alarm system on the house	A piano	A laptop
	New Zealand	Pay television (e.g. Sky, Saturn)	Do you and your family have a holiday away from home for at least one week each year? (Yes/No)	Do your parents own a holiday home?
	Norway	iPad®	iPhone	N/A
	Poland	Satellite or cable TV with at least 30 channels	Digital camera	Plasma or LCD TV
	Portugal	Cable TV or television by parabolic antenna	Plasma or LCD television	Air conditioning
	Slovak Republic	Videocamera	Digital camera (not as a part of a mobile phone, but separate one)	Lawn-mover
	Slovenia	Your own computer	Attending an extra out-of-school-time activities paid by your parents	Travelling abroad for one week or more.
	Spain	Video camera	Pay television	Home cinema
	Sweden	Piano	Jacuzzi	Espresso machine
	Switzerland	Musical instrument (excluding Recorder)	An iPhone®	A digital video camera
	Turkey	Air-conditioned type heating and cooling system	Video camera	Home theatre system
	United Kingdom (England, Wales and NI)	A premium TV package (e.g. Sky Movies, Sky Sports)	A high-definition (HD) TV	A tablet computer (e.g. iPad®)
	United Kingdom (Scotland)	A premium TV package (e.g. Sky Movies, Sky Sports)	A tablet computer (e.g. iPad®)	A musical instrument (e.g. piano, violin)
	United States	A guest room	A high-speed Internet connection	A musical instrument

[Part 2/2]

Table E.1 National household possession items

		ST26Q15	ST26Q16	ST26Q17
Partners	**Albania**	Microwave	Cultural television programs with payment	Digital camera
	Argentina	Air conditioning	LCD/Plasma TV	Washing machine
	Brazil	Cable TV	Video game	iPod®
	Bulgaria	Smart phone	Digital camera	Air-conditioner
	Colombia	Digital camera	Cable TV or Direct to Home TV	Encyclopedia
	Costa Rica	Cable TV	A console of video games	A home theatre set
	Croatia	Plasma or LCD TV	Playstation® 3	Air conditioner
	Cyprus[1, 2]	Home cinema	Cable TV	Home security alarm system
	Hong Kong-China	Plasma TV/LCD TV (40″ or above)	Piano	Pay TV Channel
	Indonesia	Digital camera	Motorcycle	Car
	Jordan	Central heating	Plasma TV set	Digital camera
	Kazakhstan	Digital fotocamera	Videocamera	Satellite antenna
	Latvia	MP3 player	Bicyle	Digital photocamera
	Liechtenstein	Musical instrument (without recorder)	An iPhone	A digital videocamera
	Lithuania	Digital camera	Press. Subscription edition (newspaper, magazine)	Cinecamera
	Macao-China	LCD television	Digital camera	iPhone®
	Malaysia	Television	Refrigerator	Air conditioner
	Montenegro	Cable TV	Plasma TV	Digital camera
	Peru	Stereo	Refrigerator	Washing machine
	Qatar	MP3 walkman	Digital video camera	Video games console
	Romania	A digital video camera	iPod®	A Home cinema system
	Russian Federation	Digital camera or video camera	Home cinema	Satellite antenna
	Serbia	Camera	Dryer	LCD TV
	Shanghai-China	Vacuum collecter	Digital camera or digital vedio recorder	Juice extractor
	Singapore	Cable television	Air conditioner	Domestic helper (e.g. full/part-time maid)
	Chinese Taipei	Piano, violin	iPod®	Digital camera
	Thailand	Air conditionning	Washing machine	Microwave Oven
	Tunisia	Flat-faced screen (TV)	Digital camera	Washing machine
	United Arab Emirates	A laptop of your own	Electronic games (Wii®, Xbox®)	iPad®
	Uruguay	Cable TV	Freezer	Laptop (XO Ceibal not included)
	Viet Nam	Air-conditioner	Motorbike	Car

1. Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2. Footnote by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

ANNEX F – TECHNICAL STANDARDS FOR PISA 2012

INTRODUCTION

The purpose of this annex is to list the set of standards upon which the PISA 2012 data collection activities will be based, as was the case for previous PISA assessments [EDU/PISA/GB(2007)4/REV1]. In following the procedures specified in the standards, the partners involved in the data collection activities contribute to creating an international dataset of a quality that allows for valid cross-national inferences to be made.

The standards for data collection and submission were developed with three major, and inter-related, goals in mind: consistency, precision and generalisability of the data. Furthermore, the standards serve to ensure a timely progression of the project in general.

- **Consistency**: Data should be collected in an equivalent fashion in all countries, using equivalent test materials. A comparable sample of the student population should perform under test conditions that are as similar as possible. Given consistent data collection (and sufficiently high response rates), test results are comparable across regions and countries. The test results in different countries will reflect differences in the performance of the students measured, and will not be caused by factors which are un-related to performance.
- **Precision**: Data collection and submission practices should leave as little room as possible for spurious variation or error. This holds for both systematic and random error sources, e.g. when the testing environment differs from one group of students to another, or when data entry procedures leave room for interpretation. An increase in precision relates directly to the quality of results one can expect: The more precise the data, the more powerful the (statistical) analyses, and the more trustworthy the results to be obtained.
- **Generalisability**: Data are collected from specific individuals, in a specific situation, and at a certain point in time. Individuals to be tested, test materials and tasks, etc. should be selected in a way that will ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected but hold for a variety of settings and are valid in the target population at large. Thus, collecting data from a representative sample of the population, for example, will lead to results that accurately reflect the level of literacy of fifteen-year-old students in a country.
- **Timeliness**: Consistency, precision and generalisability of the data can be obtained in a variety of ways. However, the tight timelines and budgets in PISA, as well as the sheer number of participating countries, preclude the option of developing and monitoring local solutions to be harmonised at a later stage in the project. Therefore, the standards specify one clear-cut path along which data collection and data submission should progress.

This document strives to establish a collective agreement of mutual accountability among countries, and of the international contractor towards the countries. This document details each standard, its rationale, and the quality assurance data that need to be collected to demonstrate that the standard has been met.

Where standards have been fully met, data will be recommended for inclusion in the PISA 2012 dataset. Where standards have not been fully met, an adjudication process will determine the extent to which the quality and international comparability of the data have been affected. The result of data adjudication will determine whether the data will be recommended for inclusion in the PISA 2012 dataset.

Since attaining the various standards is cumulative and potentially interactive (i.e. not attaining standard X is NOT the same as not attaining standards X, Y and Z), in principle each dataset should be evaluated against all standards jointly. Also, it is possible that countries' proposed plans for implementation are not, for various and often unforeseen circumstances, actually implemented (e.g. national teacher strike affecting not only response rates but also testing conditions; unforeseen National Centre budget cuts which impact on print and data management quality). Therefore, the final evaluation of standards needs to be made with respect to the data as submitted since this is the definitive indication of what may appear in the released international dataset.

If any issues with attaining standards are identified, the International Project Director initiates communication with the National Centre as soon as possible. Priority in communication rectifies the identified issues.

The PISA standards act as a benchmark of best practice. As such, the standards are designed to assist National Centres and international contractor by explicitly indicating the expectations of data quality and study implementation endorsed by the PISA Governing Board, and by clarifying the timelines of the activities involved. The standards formulate levels of

attainment, while timelines and feedback schedules of both the participating countries and the international contractor are defined in the *PISA Operations Manuals*.

As specified in the contracts for the implementation of the fifth cycle of the OECD Programme for International Student Assessment, the international contractor takes responsibility for developing and implementing procedures for assuring data quality. Therefore, the international contractor mediates, and monitors the countries' activities specified in this document, while the international contractor's adherence to the standards is monitored by the participating countries via the OECD Secretariat.

Where the Technical Standards stipulate that variations from the standards require agreement between participating countries and the Consortium, National Project Managers are asked to initiate the process of negotiation and to undertake everything possible to facilitate an agreement. Where agreement between National Project Managers and the Consortium cannot be reached, the OECD will adjudicate and resolve the issues. The OECD will also adjudicate any issues resulting from non-compliance with the Technical Standards that cannot be resolved between participating countries and the Consortium.

There are three types of standards in this document; each with a specific purpose:

- **Data Standards** refer to aspects of study implementation that directly concern the quality of the data or the assurance of that quality. These standards have been endorsed by the Technical Advisory Group and wherever proportions or quantities are specified (for example, response rates), these have reached through examination of research undertaken or reviewed by members of the Technical Advisory Group with the aim of minimising the effect of any potential bias in the data.
- **Management Standards** are in place to ensure that all PISA operational objectives are met in a timely and coordinated manner.
- **National Involvement Standards** reflect the expectations set out in the PISA 2012 Terms of Reference that the content of the PISA tests is established in consultation with national representatives with international content expertise. In particular, these standards ensure that the internationally developed instruments are widely examined for cross-national, cross-cultural and cross-linguistic validity and that the interests and involvement of national stakeholders are considered throughout the study.

FORMAT OF THE DOCUMENT

The standards are grouped into sections that relate to specific tasks in the PISA data collection process. For every section, a rationale is given explaining why standard setting is necessary. The standards in each section consist of three distinct elements. First, there are the **Standards** themselves that are numbered and are shown in shaded boxes. Second, there are **Notes** that provide additional information on the standards directly. The notes are listed after the standards in each section. Third, there are the **Quality Assurance** measures that will be used to assess if a standard has been met or not. These are listed at the end of each section. In addition, the standards contain words that have a defined meaning in the context of the standards. These words are shown in *italics* throughout the document and are clarified in the **Definitions** section at the end of the document, where the terms are listed alphabetically.

SCOPE

The standards in this document apply to data from *adjudicated entities* that include both *PISA participants* and *additional adjudicated entities*. The PISA Governing Board will approve the list of *adjudicated entities* to be included in a PISA cycle.

DATA STANDARDS

1. Target Population and sampling

Rationale: Meeting the standards specified in this section will ensure that in all countries, the students tested come from the same target population in every country, and are in a nearly equivalent age range. Therefore, the results obtained will not be confounded by potential age effects. Furthermore, to be able to draw conclusions that are valid for the entire population of fifteen-year-old students, a representative sample shall be selected for participation in the test. The size of this representative sample should not be too small, in order to achieve a certain precision of measurement in all countries. For this reason, minimum numbers of participating students and schools are specified.

Standard 1.1 The *PISA Desired Target Population* is *agreed upon* through negotiation between the National Project Manager and the international contractor, within the constraints imposed by the definition of the *PISA Target Population*.

Standard 1.2 Unless otherwise *agreed upon* only *PISA-Eligible students* participate in the test.

Standard 1.3 Unless otherwise *agreed upon, the testing period*:

- is no longer than six consecutive weeks in duration;
- does not coincide with the first six weeks of the academic year; and
- begins exactly three years from the beginning of the *testing period* in the previous PISA cycle.

Standard 1.4 Schools are sampled using *agreed upon*, established and professionally recognised principles of scientific sampling.

Standard 1.5 Students are sampled using *agreed upon*, established and professionally recognised principles of scientific sampling and in a way that represents the full population of *PISA-Eligible students*.

Standard 1.6 The *PISA Defined Target Population* covers 95% or more of the *PISA Desired Target Population*. That is, *school-level exclusions* and *within-school exclusions* combined do not exceed 5%.

Standard 1.7 The student sample size is a minimum of 4 500 assessed students for *PISA participants* and 1 500 assessed students for *additional adjudicated entities*, or the entire *PISA Defined Target Population* where the *PISA Defined Target Population* is below 4 500 and 1 500 respectively.

Standard 1.8 The school sample size is a minimum of 150 schools for *PISA participants*, and 50 schools for *additional adjudicated entities*, or all schools that have students in the *PISA Defined Target Population* where the number of schools with students in the *PISA Defined Target Population* is below 150 and 50 respectively.

Standard 1.9 The school response rate is at least 85% of sampled schools. If a response rate is below 85% then an acceptable response rate can still be achieved through *agreed upon* use of replacement schools.

Standard 1.10 The student response rate is at least 80% of all sampled students across responding schools.

Note 1.1 The Target Population and Sampling standard apply to the Main Survey but not the Field Trial.

Note 1.2 Data from schools where the student response rate is greater than 25 % will be included in the PISA dataset.

Note 1.3 For the purpose of calculating school response rates, a participating school is defined as a sampled school in which more than 50 % of sampled students respond.

Note 1.4 Guidelines for acceptable exclusions that do not affect standard adherence, are as follows:
- *School level exclusions* that are exclusions due to geographical inaccessibility, extremely small school size, administration of PISA would be not feasible within the school, and other *agreed upon* reasons and that total to less than 0.5 % of the *PISA Desired Target Population*;
- *School level exclusions* that are due to a school containing only students that would be *within-school exclusions* and that total to less than 2.0 % of the *PISA Desired Target Population*; and
- *Within-school exclusions* that total to less than 2.5 % of the *PISA Desired Target Population* – these exclusions could include, for example, students not able to do the test because of a functional disability.

Note 1.5 Principles of scientific sampling include, but are not limited to:
- The identification of appropriate stratification variables to reduce sampling variance and facilitate the computation of non-response adjustments.
- The incorporation of a *target cluster size* of 35 *PISA-Eligible students* which *upon agreement* can be increased, or reduced to a number not less than 20.

Quality assurance

- Sampling procedures as specified in the *PISA Operations Manuals*
- School sample drawn by international contractor (or if drawn by the National Centre, then verified by the international contractor)
- Student sample drawn through *KeyQuest* (or if drawn by other means, then verified by the international contractor)
- Sampling forms submitted to the international contractor
- Main Survey Review Quality Assurance Survey

2. Language of testing

Rationale: Using the language of instruction will ensure analogous testing conditions for all students within a country, thereby strengthening the consistency of the data. It is assumed that the students tested have reached a level of understanding in the language of instruction that is sufficient to be able to work on the PISA test without encountering linguistic problems (see also the criteria for excluding students from the potential assessment due to insufficient

experience in the language of assessment: *within-school exclusions*). Thus, the level of literacy in reading, mathematics and science can be assessed without interference due to a critical variation in language proficiency.

Standard 2.1 The PISA test is administered to a student in a language of instruction provided by the sampled school to that sampled student in the major domain (mathematics) of the test.

- If the language of instruction in the major domain is not well defined across the set of sampled students then, if *agreed upon*, a choice of language can be provided, with the decision being made at the student, school, or National Centre level. Agreement with the international contractor will be subject to the principle that the language options provided should be languages that are common in the community and are common languages of instruction in schools in that *adjudicated entity*.
- If the language of instruction differs across domains then, if *agreed upon,* students may be tested using test booklets in more than one language on the condition that the test language of each domain matches the language of instruction for that domain. Information obtained from the Field Trial will be used to gauge the suitability of using booklets with more than one language in the Main Survey.
- In all cases the choice of test language(s) in the test booklets is made prior to the administration of the test.

3. Field Trial participation

Rationale: The Field Trial gives countries the opportunity to try out the logistics of their test procedures and allows the international contractor to make detailed analyses of the items so that only suitable ones are included in the Main Survey.

Standard 3.1. *PISA participants* participating in the PISA 2012 Main Survey will have successfully implemented the Field Trial. Unless otherwise *agreed upon*:

- A Field Trial should occur in an assessment language if that language group represents more than 5% of the target population.
- For assessment languages that apply to between 5 and 50% of the target population, the Field Trial student sample should be a minimum of 100 students per item.
- For languages that apply to more than 50% of the target population, the Field Trial student sample should be a minimum of 200 students per item.
- For *additional adjudicated entities*, where the assessment language applies to between 5 and 100% of the target population in the entity, the Field Trial student sample should be a minimum of 100 students per item.

Note 3.1 The PISA Technical Standards for the Main Survey generally apply to the Field Trial, except for the Target Population standard, the Sampling standard, and the Quality Monitoring standard. For the Field Trial a sampling plan needs to be *agreed upon*.

Note 3.2 The Field Trial participation standard for assessment languages applicable to between 5 and 50% of the Target Population can be varied if *agreed upon*, with such agreement subject to the principle that the absence of a Field Trial for that language would not affect the Main Survey and the principle that the assessment language version is trialled in another *adjudicated entity* where the assessment language applies to more than 50% of the Target Population.

Note 3.3 The sample size for the Field Trial will be a function of the test design and will be set to achieve the standard of 200 student responses per item.

Note 3.4 Consideration will be given to reducing the required number of students per item in the Field Trial where there are fewer than 200 students in total expected to be assessed in that language in the Main Survey.

4. Adaptation of tests, questionnaires and manuals

Rationale: In order to be able to assess how the performance in a country has evolved from one PISA cycle to the other, the same instruments have to be used in the assessments. If instruments differ, then it is unclear whether changes in performance reflect changes in literacy or whether they just mirror the variation in the test items. The same holds for the assessment instruments that are used within a PISA cycle: To validly compare performance across countries, all assessment instruments have to be as similar as possible. In fact, it is of utmost importance to provide equivalent information for the students in all countries that take part in the study. Therefore, not only the assessment instruments, but also the instructions given to the students, and the procedures of data-collection have to be equivalent. To achieve this goal, other individuals who play a key role in the data-collection process, i.e. the Test Administrators, School Co-ordinators, and school associates, should receive the same information in all participating countries.

Standard 4.1	Test items used for linking are administered unchanged from their previous administration.
Standard 4.2	All test instruments are psychometrically equivalent to the *source versions*. *Agreed upon* adaptations to the local context are made if needed.
Standard 4.3	The questionnaire instruments are equivalent to the *source versions*. *Agreed upon* adaptations to the local context are made if needed.
Standard 4.4	The Test Administrator Manual and the School Co-ordinator Manual (or the School Associate Manual) are equivalent to the *source versions*. *Agreed upon* adaptations to the local context are made if needed.

Note 4.1 The quality assurance requirements for this standard apply to instruments that are in an assessment language used as a language of instruction for more than 5% of the Target Population.

Quality assurance

- *Agreed upon* Manual Adaptation Spreadsheet (MAS) and Questionnaire Adaptation Spreadsheet (QAS)
- Test Adaptation Spreadsheet (TAS) and Booklet Adaptation Spreadsheet (BAS) in which adaptations to assessment units and common booklet parts are documented. Adaptations will be checked for compliance with the PISA Translation and Adaptation Guidelines by international verifiers, and the verifiers' recommendations will be vetted by the Consortium referee.
- Verifier Reports (statistics generated by the TAS with a short qualitative report)
- Final Optical Check (FOC) Report (test booklets and questionnaires only) including key correction check
- Post-FOC review
- Field Trial and Main Survey Review Quality Assurance Surveys
- Item and scale statistics

5. Translation of tests, questionnaires and manuals

Rationale: To be able to compare the performance of students across countries, and of students with different instruction languages within a country, the linguistic equivalence of all materials is central. While Standards 4.1 to 4.4 serve to ensure that equivalent information is given to the students in all countries involved, in general, the following Standards 5.1 and 5.2 emphasise the importance of language. Again the goal is to ensure that literacy will be assessed, and not variations of information caused by differences in the translation of materials.

Standard 5.1	The following documents are translated into the assessment language in order to be linguistically equivalent to the international *source versions*. ▪ All administered test instruments ▪ All administered questionnaires ▪ The Test Administrator script from the Test Administrator (or School Associate) Manual ▪ The Coding Guides
Standard 5.2	Unless otherwise *agreed upon*, the following documents are translated/adapted into the assessment language to make them linguistically equivalent to the international *source versions*. ▪ The Test Administrator (or School Associate) Manual ▪ The School Co-ordinator (or School Associate) Manual In the case of the manuals, only *specified parts* are made linguistically equivalent.

Note 5.1 The quality assurance requirements for this standard apply to instruments that are in a language that is administered to more than 10% of the target population.

Note 5.2 The "specified parts" of manuals referred to in Standard 5.2 for which checking of the linguistic equivalence to the source versions would be undertaken are the following:
-The criteria for student eligibility
-The number of students to be sampled from each school
-The definitions, codes and instructions related to the coding of the Student Tracking Form, including examples to illustrate these codes
-The General Directions as well as instructions relating to the timing of sessions
-The Session Report Form completed by the Test Administrator for each testing session, which records session and timing information.

Quality assurance

- *Agreed upon Translation Plan* developed in accordance with the specifications in the *PISA Operations Manuals* where the *Translation Plan* would require double translation by independent translators from two *source versions*.
- *Agreed upon* Questionnaire Adaptation Spreadsheet (QAS)
- Test Adaptation Spreadsheet (TAS) and Booklet Adaptation Spreadsheet (BAS) in which adaptations to assessment units and common booklet parts are documented. Adaptations will be checked for compliance with the PISA Translation and Adaptation Guidelines by international verifiers, and the verifiers' recommendations will be vetted by the Consortium referee.
- Verifier Reports (statistics generated by the TAS and short qualitative report)
- Final Optical Check report (test booklets and questionnaires only)
- Submitted test booklets as used in the study
- Field Trial and Main Survey Review Quality Assurance Surveys
- Item and scale statistics

6. Test Administration

Rationale: Certain variations in the testing procedure are particularly likely to affect test performance. Among them are session timing, the administration of test materials and support material like rulers and calculators, the instructions given prior to testing, the rules for excluding students from the assessment etc. A full list of relevant test conditions is given in the *PISA Operations Manuals*. To ensure that the data are collected consistently, and in a comparable fashion, for all participants, it is therefore very important to keep the chain of action in the data-collection process as constant as possible.

Furthermore, the goal of the assessment is to arrive at results which cover a wide range of areas. Given the time constraints, any one student is presented only with a certain portion of the test items. Moreover, to preclude sources of random error unforeseen by the test administrators and the test designers, the students taking part in the survey have to be selected *a-priori*, in a statistically random fashion. Only then will the students participating in the study mirror the population of fifteen-year-old students in the country. The statistical analysis will take this sampling design into account, thereby arriving at results that are representative for the population at large. For these reasons, it is of utmost importance to assign the proper test booklets to the participants specified beforehand. The student tracking form is central in monitoring whether this goal has been achieved.

The Test Administrator plays a central role in all of these issues. Special consideration is therefore given to the training of the Test Administrators, ensuring that as little variation in the data as possible is caused by random or systematic variation in the activities of Test Administrators.

An important part of the testing situation relates to the relationship between Test Administrators and test participants. Therefore, any personal interaction between Test Administrators and students, either in the past or in the testing situation, counteracts the goal of collecting data in a consistent fashion across countries and participants. Strict objectivity of the Test Administrator, on the other hand, is instrumental in collecting data that reflect the level of literacy obtained, and that are not influenced by factors un-related to literacy. The results based on these data will be representative for the population under consideration.

Standard 6.1	All test sessions follow international procedures as specified in the *PISA Operations Manuals*, particularly the procedures that are: - relating to test session timing; - for maintaining test conditions; - for student tracking; and - for assigning booklets.
Standard 6.2	Test Administrators are trained in person unless a suitable alternative is *agreed upon*.
Standard 6.3	The relationship between Test Administrators and participating students must not compromise the credibility of the test session. In particular, the Test Administrator should not be the reading, mathematics, or science instructor of any student in the assessment sessions he or she will administer for PISA.

Note 6.1 Test Administrators should preferably not be school staff.

Note 6.2 Preferred training procedures for Test Administrators are described in the *PISA Operations Manuals*.

Quality assurance

- Test Administrator's Test Session Report Forms
- PISA Quality Monitors
- Main Survey Review Quality Assurance Survey

7. Implementation of national options

Rationale: These standards serve to ensure that for students participating both in the international and the national survey, the national instruments will not affect the data used for the international comparisons. Data are therefore collected consistently across countries, and potential effects like test fatigue, or learning effects from national test items, are precluded.

Standard 7.1	Only *national options* that are *agreed upon* between the National Centre and the international contractor are implemented.
Standard 7.2	Any *national option* instruments that are not part of the core component of PISA are administered after all the test and questionnaire instruments of the core component of PISA have been administered to students that are part of the international PISA sample.

8. Security of the material

Rationale: The goal of the PISA assessment is to measure the literacy levels in the content domains. Prior familiarisation with the test materials, or training to the test, will heavily degrade the consistency and validity of the data. In the extreme case, the results would only reflect how well participants are able to memorise the test items. In order to be able to assess the competencies obtained during schooling rather than short-term learning success, and to make valid international comparisons, confidentiality is extremely important.

Standard 8.1	PISA materials designated as secure are kept confidential at all times. Secure materials include all test materials, data, and draft materials. In particular: ▪ no-one other than approved project staff and participating students during the test session is able to access and view the test material; ▪ no-one other than approved project staff will have access to secure PISA data and embargoed material; and ▪ formal confidentiality arrangements will be in place for all approved project staff.

Quality assurance

- Security arrangements as specified in the *PISA Operations Manuals* or *agreed upon* variation
- National Centre Quality Monitoring
- Field Trial and Main Survey Review Quality Assurance Surveys

9. Quality monitoring

Rationale: To obtain valid results from the assessment, the data collected have to be of high quality, i.e. they have to be collected in a consistent, reliable and valid fashion. This goal is implemented first and foremost by the test administrators, who are seconded by the quality monitors. The quality monitors provide country-wide supervision of all data-collection activities.

Standard 9.1	PISA test administration is monitored using site visits by trained independent quality monitors.
Standard 9.2	An agreed number of site visits to observe Test Administration sessions are conducted in each PISA participating country/economy.
Standard 9.3	Test Administration sessions that are the subject of a site visit are randomly selected.

Note 9.1 A failure to meet the Quality Monitoring Standard in the Main Survey will lead to a significant lack of quality assurance data for other standards.

Note 9.2 The Quality Monitoring standards apply to the Main Survey but not to the Field Trial.

Note 9.3 The National Centre provides the international contractor the assistance required to implement the site visits effectively.

Quality assurance

- Curricula Vitae of the PISA Quality Monitor nominees forwarded by the National Project Manager to the international contractor
- PISA Quality Monitor Reports
- National Centre Quality Monitor Report

10. Printing of material

Rationale: Variations in print quality may affect data quality. When the quality of paper and print is very poor, the performance of students is influenced not only by their levels of literacy, but also by the degree to which test materials are legible. To rule out this potential source of error, and to increase the consistency and precision of the data collection, paper and print quality samples are solicited from National Centres in their first cycle of participation.

Standard 10.1	All student assessment material is printed using an agreed upon paper and print quality.
Standard 10.2	The cover page of all PISA assessment instruments used in schools contains all information as specified by the PISA Governing Board.
Standard 10.3	The layout and pagination of all test material is the same as in the *source versions*, *unless otherwise agreed upon*.
Standard 10.4	The layout and formatting of the questionnaire material is equivalent to the *source versions*.

Note 10.1 For National Centres that have participated in previous cycles, PISA instruments used in previous cycles or from the Field Trial preceding the Main Survey that have been submitted to the international contractor can be used for the purpose of agreeing on printing quality where the National Centre indicates that printing and paper of the same standard will be used. Otherwise, National Centres will submit a sample of printed material to the international contractor for agreement, including the cover and selected items as specified in the *PISA operations manuals.*

Note 10.2 The cover page of all PISA assessment instruments used in schools should contain all information necessary to identify the material as being part of the data-collection process for PISA, and for checking whether the data collection follows the assessment design, i.e. whether the mapping of the student on the one hand, and test booklets and questionnaires, on the other, have been correctly established. The features of the cover page referred to in Standard 10.2 are specified in the *PISA Operations Manuals.*

Quality assurance

- Submitted sample or agreement that quality will be similar to previous cycle or Field Trial versions
- Booklets submitted to international contractor to meet Standard 17.4
- Booklets submitted for The International Coding Review (ICR) (Main Survey only)
- Field Trial and Main Survey Review Quality Assurance Surveys

11. Response coding[1]

Rationale: To ensure the comparability of the data, the responses from all test participants in all participating countries have to be coded following one single coding scheme. Therefore, all coding procedures have to be standardised, and coders have to complete training sessions to master this task.

Standard 11.1	The coding scheme described in the coding guide in the distributed items is implemented according to instructions from the international contractor's item developers.
Standard 11.2	Representatives from each National Centre attend the international PISA coder training session for both the Field Trial and the Main Survey.
Standard 11.3	Both the single and multiple coding procedures as specified in the *PISA Operations Manuals* (see Note 11.1), or an *agreed upon* variation thereof, are implemented.
Standard 11.4	Coders are recruited and trained following *agreed procedures*.

Note 11.1 Preferred procedures for recruiting and training coders are outlined in the *PISA Operations Manuals*

Note 11.2 The optimum number of Coder Training session participants would depend on factors such as the expertise of National Centre staff, and resource availability.

---

[1] The terms coding, coders and codes are used instead of other terms such as marking, markers, marks, rating and raters.

Quality assurance

- Indices of inter-coder agreement
- International Coding Review (ICR)
- Coding of control scripts from previous cycle's ICR
- Field Trial and Main Survey Review Quality Assurance Surveys

12. Data submission

Rationale: The timely progression of the project, within the tight timelines given depends on the quick and efficient submission of all collected data. Therefore, one single data submission format is proposed, and countries are asked to submit only one database to the international contractor. Furthermore, to avoid potential errors when consolidating the national databases, any changes in format that were implemented subsequent to the general agreement have to be announced.

Standard 12.1	Each *PISA participant* submits its data in a single database, unless otherwise *agreed upon*.
Standard 12.2	Data are submitted in the *KeyQuest* format.
Standard 12.3	Data for all instruments are submitted. This includes the test data, questionnaire data, and tracking data as described in the *PISA Operations Manuals*.
Standard 12.4	Unless *agreed upon*, all data are submitted without recoding any of the original response variables.
Standard 12.5	Each PISA participating country's database is submitted with full documentation as specified in the *PISA Operations Manuals*.

MANAGEMENT STANDARDS

13. Communication with the international contractor

Rationale: Given the tight schedule of the project, delays in communication between the National Centres and the international contractor should be minimised. Therefore, National Centres need continuous access to the resources provided by the international contractor.

Standard 13.1	The international contractors ensure that qualified staff are available to respond to requests by the National Centres during all stages of the project. The qualified staff: ▪ are authorised to respond to National Centre queries; ▪ acknowledge receipt of National Centre queries within one working day; ▪ respond to coder queries from National Centres within one working day; and ▪ respond to other queries from National Centres within five working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.

Note 13.1 Response timelines and feedback schedules for the National Centres and the international contractor are further specified in the *PISA Operations Manuals*.

14. Notification of international and national options

Rationale: Given the tight timelines, the deadlines given in the following two standards will enable the international contractor to progress with work on time.

Standard 14.1	*National options* are *agreed upon* before 1 December in the year preceding the Field Trial and before 1 December in the year preceding the Main Survey (Standard 7.1).
Standard 14.2	The National Centre notifies the international contractor of its intention to participate in specific international options before 1 December in the year preceding the Field Trial.

15. Schedule for submission of materials

Rationale: To meet the requirements of the work programme, and to progress according to the timelines of the project, the international contractor will need to receive a number of materials on time.

Standard 15.1 An *agreed upon Translation Plan* and *Preferred Verification Schedule* will be negotiated between each National Centre and the international contractor.

Standard 15.2 The following items are submitted to the international contractor in accordance with *agreed timelines*:

- the Translation Plan and Preferred Verification Schedule;
- a print sample of booklets prior to final printing (where this is required, see Standard 10.1 and Note 10.1);
- sampling forms (see Standard 1);
- Demographic Tables;
- Field Trial and Main Survey Reviews; and
- other documents as specified in the *PISA Operations Manuals*.

Standard 15.3 Questionnaire materials are submitted for linguistic verification only after all adaptations have been *agreed upon*.

Standard 15.4 Those elements of the Test Administrator and School Co-ordinator (or School Associate) manuals requiring verification, including linguistic verification as specified in Standard 5.2, are submitted only after all adaptations have been *agreed upon*.

Quality assurance

- Agreed upon Translation Plan and Preferred Verification Schedule
- International contractor records
- Test materials are submitted for linguistic verification with corresponding adaptation spreadsheets filled in by the National Centre

16. Drawing samples

Rationale: The mode of drawing the samples used in the study is crucial to data quality. The goal of the project is to collect data that are representative for the population at large. To reach this goal, the sampling procedures have to follow established scientific rules. Furthermore, the comparability of the data across countries is guaranteed if the same procedure is used for all national samples. If different sampling procedures are used, then the equivalence of the sampling quality has to be determined.

Standard 16.1 For efficient and effective quality assurance provision, unless otherwise *agreed upon*, the international contractor will draw the school sample for the Main Survey.

Agreement with the international contractor will be subject to the principle that the sampling methods used are scientifically valid and consistent with PISA's documented sampling methods. Where a *PISA participating country* chooses to draw the school sample, the National Centre provides the international contractor with the data and documentation required for it to verify the correctness of the sampling procedures applied.

Standard 16.2 For efficient and effective quality assurance provision, unless otherwise *agreed upon*, the National Centre will use *KeyQuest* to draw the student sample, using the list of eligible students provided for each school.

Where a *PISA participating country* chooses not to use *KeyQuest* to draw the student sample, the National Centre provides the international contractor with the data and documentation required for it to verify the correctness of the sampling procedures applied.

Note 16.1 Any costs associated with verifying a school sample taken by the National Centre, or a student sample selected other than by using *KeyQuest* will be borne by the National Centre.

17. Management of data

Rationale: Consolidating and merging the national databases is a time-consuming and difficult task. To ensure the timely and efficient progress of the project, the international contractor needs continuous access to national resources helping to rule out uncertainties and to resolve discrepancies. This standard aims to prevent substantial delays to the whole project which could result from a delay in processing the data of a small number of participating countries.

Standard 17.1	The timeline for submission of national databases to the international contractor is within eight weeks of the last day of testing for the Field Trial and within twelve weeks of the last day of testing for the Main Survey, unless otherwise *agreed upon*.
Standard 17.2	National Centres execute data checking procedures as specified in the *PISA Operation Manuals* before submitting the database.
Standard 17.3	National Centres make a data manager available upon submission of the database. The data manager: ▪ is authorised to respond to international contractor data queries; ▪ is available for a three-month period immediately after the database is submitted unless otherwise *agreed upon*; ▪ is able to respond to international contractor queries within three working days; and ▪ is able to resolve data discrepancies.
Standard 17.4	A complete set of PISA instruments as administered and including any *national options*, is forwarded to the international contractor on or before the first day of testing. The submission includes the following: ▪ hard copies of instruments; and ▪ electronic PDF copies of instruments
Standard 17.5	To enable the *PISA participant* to submit a single dataset, all instruments for all *additional adjudicated entities* will contain the same variables as the primary *adjudicated entity* of the *PISA participan*t.

Note 17.1 Each participating country/economy will receive its own national micro-level PISA database (the “national database”), in electronic form as soon as it has been processed from the international contractor for PISA. The national database will contain the complete set of responses from the students, parents, school principals and surveyed participants in that country/economy.

Each participating country/economy has access to and can publish its own data after a date that is established by the PISA Governing Board for the publication of the initial OECD publication of the survey results (the “initial international OECD publication”).

The OECD Secretariat will not release national data to other countries/economies until participating countries/economies have been given an opportunity to review and comment on their own national data and until the release of such data has been approved by the national authorities.

A deadline and procedures for withdrawing countries/economies’ national data from the international micro-level PISA database (the “international database”) will be decided upon by the PISA Governing Board. Countries/economies can withdraw data only prior to obtaining access to data from other countries/economies. Withdrawn data will not be made available to other countries/economies.

The PISA Governing Board will discuss with participating countries/economies whose data manifests technical anomalies as to whether the data concerned can be included in the international database. The decision of the PISA Governing Board will be final. Participating countries/economies may, however, continue to use data that are excluded from the international database at the national level.

The OECD Secretariat will then compile the international database, which will comprise the complete set of national PISA databases, except those data elements that have been withdrawn by participating countries/economies or by the PISA Governing Board at the previous stage. The international database will remain confidential until the date on which the initial international OECD publication is released.

National data from all participating countries/economies represented in the international database will be made available to all participating countries/economies from the date on which the initial international OECD publication is released.

After release of the initial international OECD publication, the international database will be made publicly available on a cost-free basis, through the OECD Secretariat. The database may not be offered for sale.

The international database will form the basis for OECD indicator reports and publications.

The international contractor for PISA 2012 will have no ownership of instruments or data nor any rights of publication and will be subject to the confidentiality terms set in this agreement.

The OECD establishes rules to ensure adherence to the above procedure and to the continued confidentiality of the PISA data and materials until the agreed release dates. These include confidentiality agreements with all individuals that have access to the PISA material prior to its release.

As guardian of the process and producer of the international database, the OECD will hold copyright in the database and in all original material used to develop, or be included in, the PISA Field Trial and PISA Main Survey (among them the assessment materials, field manuals, and coding guides) in any language and format.

Quality assurance

- International contractor Records

18. Archiving of materials

Rationale: The international contractor will maintain an electronic archive. This will provide an overview of all materials used and ensure continuity of materials available in participating countries across PISA survey cycles, therefore building upon the knowledge gained nationally in the course of the PISA cycles. This will also ensure that the international contractor has the relevant materials available during data cleaning, when they are first required.

Standard 18.1 The international contractor will maintain a permanent electronic archive of all assessment materials, field manuals and coding guides. To facilitate this, the National Project Manager submits one copy of each of the following translated and adapted Main Survey materials to the international contractor in the source version software format:

- all administered Test Instruments, including *national options*;
- all administered Questionnaires, including *national options*;
- Test Administrator, School Co-ordinator and School Associate manuals; and
- Coding Guides.

Standard 18.2 Unless otherwise requested, National Centres will archive all Field Trial materials until the beginning of the Main Survey, and all Main Survey materials until the publication of the international report. Materials to be archived include:

- all respondents' test booklets and questionnaires;
- sampling forms;
- student lists;
- student tracking instruments; and
- all data submitted to the international contractor.

After completion of a survey the National Centre will transfer this archive to international contractor who will compile the national archives from all participants and transfer them to OECD after completion of the Main Survey.

NATIONAL INVOLVEMENT STANDARDS

19. National feedback

National feedback in areas such as test development is important in maintaining the dynamic and collaborative nature of PISA. National feedback ensures that instruments achieve cross-national, cross-cultural and cross-linguistic validity. It also promotes the inclusion of the interests and involvement of national stakeholders.

Standard 19.1 National Centres develop appropriate mechanisms in order to promote participation, effective implementation, and dissemination of results amongst all relevant national stakeholders.

Standard 19.2 National Centres provide feedback to the international contractor on the development of instruments, domain frameworks, the adaptation of instruments, and other domain related matters that represents the perspectives of the relevant national stakeholders.

Note 19.1 As a guideline feedback might be sought from the following relevant stakeholders: policy makers, curriculum developers, domain experts, test developers, linguistic experts and experienced teachers.

Quality assurance

- National Centre Quality Monitoring
- Documented strategies
- List of committees and groups
- Membership records of representative groups and/or committees
- Meeting records of representative groups and/or committees

DEFINITIONS

Additional Adjudicated Entities – entities in addition to the first and primary entity managed by a *PISA participant*, where a *PISA participant* manages more than one *adjudicated entity*.

Adjudicated Entity – a country, geographic region, or similarly defined population, for which the international contractor fully implements quality assurance and quality control mechanisms and endorses, or otherwise, the publication of separate PISA results.

Agreed procedures – procedures that are specified in the *PISA Operations Manuals*, or variations that are *agreed upon* between the National Project Manager and the international contractor.

Agreed timelines – timelines that are specified in the *PISA Operations Manuals*, or variations that are *agreed upon* between the National Project Manager and the international contractor.

Agreed upon – variations and definitions agreed upon between the National Project Manager and the international contractor. Agreed upon variations are available to National Project Managers on their National Centre webpage on the *international contractor Website*.

International contractor website – website with address *http://mypisa.acer.edu.au*. This website contains the *source versions* of instruments, manuals and other documents and information relating to National Centres. These materials are also available from *www.oecd.org/pisa*.

International Coding Review – a quality assurance exercise that requires National Centres to send a sample of student test booklets to the international contractor. The booklets required for the quality assurance study will be identified by the international contractor after the National Centre's data has been submitted. The number of booklets to be submitted by each *PISA participating country/economy* will depend on the number of languages of assessment, the number of adjudicated entities, and the number of coding centres used.

International Option – optional additional international instruments or procedures designed and fully supported by the international contractor.

***KeyQuest* –** software developed by the international contractor specifically for the PISA project. The software assists with sampling, student tracking and data submission practices that meet the PISA 2012 technical standards.

National Centre Quality Monitor – an international contractor representative who visits a National Centre in the month preceding the Main Survey to train *PISA Quality Monitors* and conduct a scheduled interview with the National Project Manager.

National Option – A *national option* occurs if:

i) A National Centre administers any additional instrumentation, for example a test or questionnaire, to schools or students that are part of the PISA international sample. Note that in the case of adding items to the questionnaires, an addition of five or more items to either the school questionnaire or the student questionnaire is regarded as a *national option*.

OR

ii) A National Centre administers any PISA international instrumentation to any students or schools that are not part of an international PISA sample (age-based or grade-based) and therefore will not be included in the respective PISA international database.

PISA Defined Target Population – all *PISA-Eligible students* in the schools that are listed on the school sampling frame. That is, the *PISA Desired Target Population* minus exclusions.

PISA Desired Target Population – the *PISA Target Population* defined for a specific *adjudicated entity*. It provides the most exhaustive coverage of *PISA-Eligible students* in the *adjudicated entity* as is feasible.

PISA-Eligible Students – students who are in the *PISA Target Population*.

PISA Operations Manuals – manuals provided by the international contractor, that is the following:

- National Project Manager's Manual;
- Test Administrator Manual;
- School Coordinator Manual;
- School Associate Manual;
- School Sampling Preparations Manual;
- Data Management Manual; and
- all other key documents referenced within the National Project Manager's manual.

The preparation of the *PISA Operations Manuals* will be carried out by the international contractor and will describe procedures developed by the international contractor. The manuals will be prepared following consultation with participating countries/economies, the OECD Secretariat, the Technical Advisory Group and other stakeholders.

PISA Participant – an administration centre, commonly called a National Centre that is managed by a person, commonly called a National Project Manager, who is responsible for administering PISA in an *adjudicated entity* and in zero or more *additional adjudicated entities*. The National Project Manager must be authorised to communicate with the international contractor on all operational matters relating to the *adjudicated entities* for which the National Project Manager is responsible.

PISA Quality Monitor – a person nominated by the National Project Manager and employed by the international contractor to monitor Test Administration quality in an adjudicated entity.

PISA Target Population – students aged between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the *testing period,* attending educational institutions located within the *adjudicated entity*, and in grade 7 or higher. The age range of the population may vary up to one month, either older or younger, but the age range must remain 12 months in length. That is, the population can be as young as between 15 years and 2 (completed) months and 16 years and 1 (completed) month at the beginning of the *testing period;* or as old as between 15 years and 4 (completed) months and 16 years and 3 (completed) months at the beginning of the *testing period*.

Preferred Verification Schedule – a schedule that provides a timeline for the submission of material relating to the adaptation of instruments and the submission of instruments for linguistic verification including the Final Optical Check. This schedule can be found in the PISA National Project Manager's Manual.

School Level Exclusions – exclusion of schools from the sampling frame because:

- of geographical inaccessibility (but not part of a region that is omitted from the *PISA Desired Target Population*);
- of an extremely small size;
- administration of the PISA assessment within the school would not be feasible;
- all students in the school would be *within-school exclusions*; or
- of other reasons as *agreed upon*.

Source Versions – documents provided in English and French by the international contractor.

Target Cluster Size – the number of students that are to be sampled from schools where not all students are to be included in the sample.

Testing Period – the period of time during which data is collected in an *adjudicated entity*.

Translation Plan – documentation of all the processes that are intended to be used for all activities related to translation and languages.

Within-school exclusions – exclusion of students from potential assessment because of one of the following:

- They are functionally disabled in such a way that they cannot take the PISA test. Functionally disabled students are those with a moderate to severe permanent physical disability.
- They have a cognitive, behavioural or emotional disability confirmed by qualified staff, meaning they cannot take the PISA test. These are students who are cognitively, behaviourally or emotionally unable to follow even the general instructions of the assessment.
- They have insufficient assessment language experience to take the PISA test. Students who have insufficient assessment language experience are those who meet all the following three criteria:
  - they are not native speakers of the assessment language;
  - they have limited proficiency in the assessment language; and
  - they have received less than one year of instruction in the assessment language.
- They cannot be assessed for some other reason as *agreed upon*.

ANNEX G – PISA CONSORTIUM, STAFF AND CONSULTANTS

PISA is a collaborative effort, bringing together experts from the participating countries, steered jointly by their governments on the basis of shared, policy-driven interests.

A PISA Governing Board, on which each country is represented, determines the policy priorities for PISA, in the context of OECD objectives, and oversees adherence to these priorities during the implementation of the programme. This includes setting priorities for the development of indicators, for establishing the assessment instruments, and for reporting the results.

Experts from participating countries also serve on working groups that are charged with linking policy objectives with the best internationally available technical expertise. By participating in these expert groups, countries ensure that the instruments are internationally valid and take into account the cultural and educational contexts in OECD member and partner countries and economies, that the assessment materials have strong measurement properties, and that the instruments place emphasise authenticity and educational validity.

Through National Project Managers, participating countries and economies implement PISA at the national level subject to the agreed administration procedures. National Project Managers play a vital role in ensuring that the implementation of the survey is of high quality, and verify and evaluate the survey results, analyses, reports and publications.

The design and implementation of the surveys, within the framework established by the PISA Governing Board, is the responsibility of external contractors. For PISA 2012, the development and implementation of the cognitive assessment and questionnaires, and of the international options, was carried out by a consortium led by the Australian Council for Educational Research (ACER). Other partners in this Consortium include cApStAn Linguistic Quality Control in Belgium, the Centre de Recherche Public Henri Tudor (CRP-HT) in Luxembourg, the Department of Teacher Education and School Research (ILS) at the University of Oslo in Norway, the Deutsches Institut für Internationale Pädagogische Forschung (DIPF) in Germany, the Educational Testing Service (ETS) in the United States, the Leibniz Institute for Science and Mathematics Education (IPN) in Germany, the National Institute for Educational Policy Research in Japan (NIER), the Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe) at the University of Liège in Belgium, and WESTAT in the United States, as well as individual consultants from several countries. ACER also collaborated with Achieve, Inc. in the United States to develop the mathematics framework for PISA 2012.

The OECD Secretariat has overall managerial responsibility for the programme, monitors its implementation daily, acts as the secretariat for the PISA Governing Board, builds consensus among countries and serves as the interlocutor between the PISA Governing Board and the international Consortium charged with implementing the activities. The OECD Secretariat also produces the indicators and analyses and prepares the international reports and publications in co-operation with the PISA Consortium and in close consultation with member and partner countries and economies both at the policy level (PISA Governing Board) and at the level of implementation (National Project Managers).

PISA Governing Board

Chair of the PISA Governing Board: Lorna Bertrand

OECD countries

Australia: Tony Zanderigo
Austria: Mark Német
Belgium: Christiane Blondin and Isabelle Erauw
Canada: Pierre Brochu, Patrick Bussiere and Tomasz Gluszynski
Chile: Leonor Cariola Huerta
Czech Republic: Jana Paleckova
Denmark: Tine Bak and Elsebeth Aller
Estonia: Maie Kitsing
Finland: Tommi Karjalainen
France: Bruno Trosseille
Germany: Elfriede Ohrnberger and Susanne von Below
Greece: Vassilia Hatzinikita and Chryssa Sofianopoulou
Hungary: Benõ Csapó
Iceland: Júlíus Björnsson
Ireland: Jude Cosgrove and Gerry Shiel

Israel: Michal Beller and Hagit Glickman
Italy: Paolo Sestito
Japan: Ryo Watanabe
Korea: Sungsook Kim and Keunwoo Lee
Luxembourg: Amina Kafai
Mexico: Francisco Ciscomani and Eduardo Backhoff Escudero
Netherlands: Paul van Oijen
New Zealand: Lynne Whitney
Norway : Anne-Berit Kavli and Alette Schreiner
Poland: Stanislaw Drzazdzewski and Hania Bouacid
Portugal: Luisa Canto and Castro Loura
Slovak Republic: Romana Kanovska and Paulina Korsnakova
Slovenia: Andreja Barle Lakota
Spain: Ismael Sanz Labrador
Sweden: Anita Wester
Switzerland: Vera Husfeldt and Claudia Zahner Rossier
Turkey: Nurcan Devici and Mustafa Nadir Çalis
United Kingdom: Lorna Bertrand and Jonathan Wright
United States: Jack Buckley, Dana Kelly and Daniel McGrath

Observers

Albania: Ermal Elezi
Argentina: Liliana Pascual
Brazil: Luiz Claudio Costa
Bulgaria: Neda Kristanova
Colombia: Adriana Molina
Costa Rica: Leonardo Garnier Rimolo
Croatia: Michelle Bras Roth
Hong Kong-China: Esther Sui-chu Ho
Indonesia: Khairil Anwar Notodiputro
Jordan: Khattab Mohammad Abulibdeh
Kazakhstan: Almagul Kultumanova
Latvia: Andris Kangro, Ennata Kivrina and Dita Traidas
Lithuania: Rita Dukynaite
Macao-China: Leong Lai
Montenegro: Zeljko Jacimovic
Panama: Arturo Rivera
Peru: Liliana Miranda Molina
Qatar: Hamda Al Sulaiti
Romania: Roxana Mihail
Russian Federation: Isak Froumin and Galina Kovaleva
Serbia: Dragica Pavlovic-Babic
Shanghai-China: Minxuan Zhang
Singapore: Khah Gek Low
Chinese Taipei: Gwo-Dong Chen and Chih-Wei Hue
Thailand: Precharn Dechsri
United Arab Emirates: Moza al Ghufly and Ayesha G. Khalfan Almerri
Uruguay: Andrés Peri and Maria Helvecia Sanchez Nunez
Viet Nam: Le Thi My Ha

PISA 2012 National Project Managers

Albania: Alfonso Harizaj
Argentina: Liliana Pascual
Australia: Sue Thomson
Austria: Ursula Schwantner
Belgium: Inge De Meyer and Ariane Baye
Brazil: João Galvão Bacchetto
Bulgaria: Svetla Petrova
Canada: Pierre Brochu and Tamara Knighton
Chile: Ema Lagos Campos
Colombia: Francisco Reyes
Costa Rica: Lilliam Mora
Croatia: Michelle Bras Roth
Czech Republic: Jana Paleckova
Denmark: Niels Egelund
Estonia: Gunda Tire

Finland: Jouni Välijärvi
France: Ginette Bourny
Germany: Christine Sälzer and Manfred Prenzel
Greece: Vassilia Hatzinikita
Hong Kong-China: Esther Sui-chu Ho
Hungary: Ildikó Balazsi
Iceland: Almar Midvik Halldorsson
Indonesia: Yulia Wardhani Nugaan and Hari Setiadi
Ireland: Gerry Shiel and Rachel Perkins
Israel: Joel Rapp and Inbal Ron-Kaplan
Italy: Carlo Di Chiacchio
Japan: Ryo Watanabe
Jordan: Khattab Mohammad Abulibdeh
Kazakhstan: Gulmira Berdibayeva and Zhannur Azmagambetova
Korea: Ji-Min Cho and Mi-Young Song
Latvia: Andris Kangro
Liechtenstein: Christian Nidegger
Lithuania: Mindaugas Stundza
Luxembourg: Bettina Boehm
Macao-China: Kwok Cheung Cheung
Malaysia: Ihsan Ismail and Muhamad Zaini Md Zain
Mexico: María Antonieta Díaz Gutierrez
Montenegro: Divna Paljevic Sturm
Netherlands: Jesse Koops
New Zealand: Kate Lang and Steven May
Norway: Marit Kjaernsli
Peru: Liliana Miranda Molina
Poland: Michal Federowicz
Portugal: Ana Sousa Ferreira
Qatar: Aysha Al-Hashemi and Assad Tounakti
Romania: Silviu Cristian Mirescu
Russian Federation: Galina Kovaleva
Scotland: Rebecca Wheater
Serbia: Dragica Pavlovic-Babic
Shanghai-China: Jing Lu and Minxuan Zhang
Singapore: Chew Leng Poon and Sean Tan
Slovak Republic: Julia Miklovicova and Jana Ferencova
Slovenia: Mojca Straus
Spain: Lis Cercadillo Pérez
Sweden: Magnus Oskarsson
Switzerland: Christian Nidegger
Chinese Taipei: Pi-Hsia Hung
Thailand: Sunee Klainin
Tunisia: Mohamed Kamel Essid
Turkey: Serdar Aztekin
United Arab Emirates: Moza al Ghufly
United Kingdom: Rebecca Wheater
United States: Dana Kelly and Holly Xie
Uruguay: Maria Helvecia Sánchez Nunez
Viet Nam: Thi My Ha Le

OECD Secretariat

Andreas Schleicher (Strategic development)
Francesco Avvisati (Analytic services)
Simone Bloem (Analytic services)
Francesca Borgonovi (Analytic services)
Jenny Bradshaw (Project management)
Claire Chetcuti (Administrative support)
Michael Davidson (Project management and analytic services)
Cassandra Davis (Dissemination co-ordination)
Juliet Evans (Administration and partner country/economy relations)
Miyako Ikeda (Analytic services)
Maciej Jakubowski (Consultant, technical review)
Guillermo Montt (Analytic services)
Josefa Palacios (Analytic services)
Giannina Rech (Analytic services)

Sophie Vayssettes (Project management and analytic services)
Elisabeth Villoutreix (Production co-ordination and editorial support)
Pablo Zoido (Analytic services)

PISA 2012 mathematics expert group

Kaye Stacey (Chair) (University of Melbourne, Australia)
Caroline Bardini (University of Melbourne, Australia)
Werner Blum (University of Kassel, Germany)
Joan Ferrini-Mundy (Michigan State University, United States)
Solomon Garfunkel (COMAP, United States)
Toshikazu Ikeda (Yokohama National University, Japan)
Zbigniew Marciniak (Warsaw University, Poland)
Mogens Niss (Roskilde University, Denmark)
Martin Ripley (World Class Arena Limited, United Kingdom)
William Schmidt (Michigan State University, United States)

PISA 2012 problem solving expert group

Joachim Funke (Chair) (University of Heidelberg, Germany)
Benő Csapó (University of Szeged, Hungary)
John Dossey (Illinois State University, United States)
Arthur Graesser (The University of Memphis United States)
Detlev Leutner (Duisburg-Essen University, Germany)
Romain Martin (Université de Luxembourg FLSHASE, Luxembourg)
Richard Mayer (University of California, United States)
Ming Ming Tan (Ministry of Education, Singapore)

PISA 2012 financial literacy expert group

Annamaria Lusardi (Chair) (The George Washington University School of Business, United States)
Jean-Pierre Boisivon (Université de Paris II Panthéon-Assas, France)
Diana Crossan (Commission for Financial Literacy and Retirement Income, New Zealand)
Peter Cuzner (Australian Securities and Investments Commission, Australia)
Jeanne Hogarth (Federal Reserve System, United States)
Dušan Hradil (Ministry of Finance, Czech Republic)
Stan Jones (Consultant, Canada)
Sue Lewis (Consultant, United Kingdom)

PISA 2012 questionnaire expert group

Eckhard Klieme (Chair) (Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Germany)
Eduardo Backhoff (University of Baja California at the Institute of Educational Research and Development, Mexico)
Ying-yi Hong (Nanyang Business School of Nanyang Technological University, Singapore)
David Kaplan (University of Wisconsin – Madison, United States)
Henry Levin (Columbia University, United States)
Jaap Scheerens (University of Twente, Netherlands)
William Schmidt (Michigan State University, United States)
Fons van de Vijver (Tilburg University, Netherlands)

Technical advisory group

Keith Rust (Chair) (Westat, United States)
Ray Adams (ACER, Australia)
Cees Glas (University of Twente, Netherlands)
John de Jong (Language Testing Services, Netherlands)
David Kaplan (University of Wisconsin – Madison, United States)
Christian Monseur (University of Liège, Belgium)
Sophia Rabe-Hesketh (University of California – Berkeley, United States)
Thierry Rocher (Ministry of Education, France)
Norman Verhelst (CITO, Netherlands)
Kentaro Yamamoto (ETS, United States)
Rebecca Zwick (University of California, United States)

PISA 2012 Consortium

Australian Council for Educational Research

Ray Adams (International Project Director)
Susan Bates (Project administration)

Alla Berezner (Data management and analysis)
Yan Bibby (Data processing and analysis)
Phillipe Bickham (IT services)
Esther Brakey (Administrative support)
Robin Buckley (IT services)
Mark Butler (Financial literacy instruments and test development)
Wei Buttress (Project administration and quality monitoring)
Renee Chow (Data processing and analysis)
John Cresswell (Reporting and dissemination)
Alex Daraganov (Data processing and analysis)
Jorge Fallas (Data processing and analysis)
Kate Fitzgerald (Data processing and sampling)
Kim Fitzgerald (IT Services)
Paul Golden (IT and helpdesk support)
Jennifer Hong (Data processing and sampling)
Nora Kovarcikova (Survey operations)
Winson Lam (IT services)
Petra Lietz (Questionnaire development)
Tom Lumley (Reading instruments and test development)
Greg Macaskill (Data management and processing and sampling)
Ron Martin (Science instruments and test development)
Barry McCrae (Problem solving and science instruments and test development)
Louise McDonald (Graphic design)
Juliette Mendelovits (Reading and financial literacy instruments and test development)
Martin Murphy (Field operations and sampling)
Thoa Nguyen (Data processing and analysis)
Stephen Oakes (IT management and support)
Elizabeth O'Grady (Questionnaire development and project support)
Penny Pearson (Administrative support)
Ray Peck (Mathematics and financial literacy instruments and test development)
Fei Peng (Quality monitoring and project support)
Ray Philpot (Problem Solving instruments and test development)
Anna Plotka (Graphic design)
Dara Ramalingam (Reading instruments and test development)
Sima Rodrigues (Data processing and analysis)
Alla Routitsky (Data management and processing)
James Spithill (Mathematics instruments and test development)
Rachel Stanyon (Project support)
Naoko Tabata (Survey operations)
Stephanie Templeton (Project administration and support)
Mollie Tobin (Questionnaire development and project support)
David Tout (Mathematics instruments and test development)
Ross Turner (Management, mathematics instruments and test development)
Maryanne Van Grunsven (Project support)
Charlotte Waters (Project administration, data processing and analysis)
Maurice Walker (Management, computer-based assessment)
Louise Wenn (Data processing and analysis)
Yan Wiwecka (IT services)

cApStAn Linguistic Quality Control (BELGIUM)

Raphael Choppinet (Computer-based verification management)
Steve Dept (Translation and verification operations)
Andrea Ferrari (Linguistic quality assurance and quality control designs)
Musab Hayatli (Right-to-left scripts, cultural adaptations)
Elica Krajceva (Questionnaire verification co-ordinator)
Shinoh Lee (Cognitive test verification co-ordinator)
Irene Liberati (Manuals verification co-ordinator)
Laura Wayrynen (Verifier training and verification procedures)

Educational Testing Service (ETS)

Jonas Bertling (Questionnaire instruments and test development)
Irwin Kirsch (Reading Components)
Patricia Klag (Problem-solving instruments and test development)
Patrick Kyllonen (Questionnaire instruments and test development)
Marylou Lennon (Questionnaire instruments and test development)
Richard Roberts (Questionnaire instruments and test development)

Matthias von Davier (Questionnaire instruments and test development)
Kentaro Yamamoto (Member TAG, problem-solving instruments and test development)

Deutches Institue für Internationale Pädagogische Forschung (DIPF, GERMANY)

Frank Goldhammer (Test developer, problem solving)
Eckhard Klieme (Chair of Questionnaire Expert Group)
Silke Hertel (Questionnaire development)
Jean-Paul Reeff (International Consultant)
Heiko Rolke (Software Design & Software Development Management [Delivery System, Translation System])
Brigitte Steinert (Questionnaire development)
Svenja Vieluf (Questionnaire development)

Institutt for Lærerutdanning Og Skoleutvikling (ILS, NORWAY)

Bjornar Alseth (Mathematics instruments and test development)
Ole Kristian Bergem (Mathematics instruments and test development)
Knut Skrindo (Mathematics instruments and test development)
Rolf V. Olsen (Mathematics instruments and test development)
Arne Hole (Mathematics instruments and test development)
Therese Hopfenbeck (Problem-solving instruments and test development)

Leibniz Institute for Science and Mathematics Education (IPN, GERMANY)

Christoph Duchhardt (Mathematics instruments and test development)
Aiso Heinze (Mathematics instruments and test development)
Eva Knopp (Mathematics instruments and test development)
Martin Senkbeil (Mathematics instruments and test development)

National Institute for Educational Policy Research (NIER, JAPAN)

Keiichi Nishimura (Mathematics instruments and test development)
Yuji Surata (Mathematics instruments and test development)

The TAO Initiative: Henry Tudor Public Research Centre, University of Luxembourg (LUXEMBOURG)

Joel Billard (Software Engineer, School Questionnaire)
Marilyn Binkley (Project Consultant, Assessment Expert)
Jerome Bogaerts (Software Engineer, TAO Platform)
Gilbert Busana (Electronic Instruments, Usability)
Christophe Henry (System Engineer, School Questionnaire and Hosting)
Raynald Jadoul (Technical Lead, School Questionnaire and Electronic Instruments)
Isabelle Jars (Project Manager)
Vincent Koenig (Electronic Instruments, Usability)
Thibaud Latour (Project Leader, TAO Platform)
Lionel Lecaque (Software Engineer, Quality)
Primael Lorbat (Software Engineer, Electronic Instruments)
Romain Martin (Problem Solving Expert Group Member)
Matteo Melis (Software Engineer, School Questionnaire)
Patrick Plichart (Software Architect, TAO Platform)
Vincent Porro (Software Engineer, Electronic Instruments)
Igor Ribassin (Software Engineer, Electronic Instruments)
Somsack Sipasseuth (Software Engineer, Electronic Instruments)

Unité d’analyse des Systèmes et des Pratiques d’enseignement (ASPE, BELGIUM)

Isabelle Demonty (Mathematics instruments and test development)
Annick Fagnant (Mathematics instruments and test development)
Anne Matoul (French source development)
Christian Monseur (Member of Technical Advisory Group)

WESTAT

Susan Fuss (Sampling and weighting)
Amita Gopinath (Weighting)
Jing Kang (Sampling and weighting)
Sheila Krawchuk (Sampling, weighting and quality monitoring)
Thanh Le (Sampling, weighting and quality monitoring)
John Lopdell (Sampling and weighting)
Keith Rust (Director of the PISA Consortium for sampling and weighting)
Erin Willey (Sampling and weighting)
Shawn Lu (Weighting)

Teresa Strickler (Weighting)
Yumiko Sugawara (Weighting)
Joel Wakesberg (Sampling and weighting)
Sergey Yagodin (Weighting)

Achieve Inc.

Michael Cohen (Mathematics framework development)
Kaye Forgione (Mathematics framework development)
Morgan Saxby (Mathematics framework development)
Laura Slover (Mathematics framework development)
Bonnie Verrico (Project support)

HallStat SPRL

Beatrice Halleux (Consultant, translation/verification referee, French source development)

University of Heidelberg

Joachim Funke (Chair, Problem Solving Expert Group)
Samuel Greiff (Problem-solving instruments and test development)

University of Melbourne

Caroline Bardini (Member Mathematics Expert Group)
John Dowsey (Mathematics instruments and test development)
Derek Holton (Mathematics instruments and test development)
Kaye Stacey (Chair Mathematics Expert Group)

Other experts

Michael Besser (Mathematics instruments and test development, University of Kassel, Germany)
Khurrem Jehangir (Data analysis for TAG, University of Twente, Netherlands)
Kees Lagerwaard (Mathematics instruments and test development, Institute for Educational Measurement of Netherlands, Netherlands)
Dominik Leiss (Mathematics instruments and test development, University of Kassel, Germany)
Anne-Laure Monnier (Consultant French source development, France)
Hanako Senuma (Mathematics instruments and test development, Tamagawa University, Japan)

Publication layout

Groupe Jouve

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16

PISA 2012
Technical Report

The *PISA 2012 Technical Report* describes the methodology underlying the PISA 2012 survey which tested 15-year-olds' competencies in mathematics, reading and science and, for some countries, problem solving and financial literacy. It examines additional features related to the implementation of the project at a level of detail that allows researchers to understand and replicate its analyses. The reader will find a wealth of information on the test and sample design, modes of administration (paper-based or computer-based), methodologies used to analyse the data, technical features of the project, and quality control mechanisms.

Contents of this volume

2014